2018

BIOSYNTHETIC MECHANISM OF THE ANTIBIOTIC CAPURAMYCIN

Erfu Yan
University of Kentucky, justinyunnan@gmail.com
Author ORCID Identifier:
https://orcid.org/0000-0002-5545-0995
Digital Object Identifier: https://doi.org/10.13023/etd.2018.360

STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Erfu Yan, Student

Dr. Steven Van Lanen, Major Professor

Dr. David Feola, Director of Graduate Studies

BIOSYNTHETIC MECHANISM OF THE ANTIBIOTIC CAPURAMYCIN

DISSERTATION

A dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy in the
College of Pharmacy
at the University of Kentucky

By
Erfu Yan

Lexington, Kentucky

Director: Dr. Steven Van Lanen, Professor of Pharmaceutical Science

Lexington, Kentucky

2018

ABSTRACT OF DISSERTATION

BIOSYNTHETIC MECHANISM OF THE ANTIBIOTIC CAPURAMYCIN

A-102395 is a member of the capuramycin family of antibiotics which was isolated from the culture broth of *Amycolatopsis sp.* SANK 60206. A-102339 is structurally classified as a nucleoside antibiotic, which like all members of the capuramycin family, inhibits bacterial MraY (translocase I) with IC_{50} of 11 nM which is the lowest among the capuramycin family. A semisynthetic derivative of capuramycin is currently in clinical trials as an antituberculosis antibiotic, suggesting high potential for using A-102395 as a starting point for new antibiotic discovery. In contrast to other capuramycins, A-102395 has a unique arylamine-containing polyamide side chain. The biosynthetic gene cluster of A-102395 was previously identified and includes 35 putative open reading frames responsible for biosynthesis and resistance. Presently, there are no reports focused on the biosynthesis of this polyamide chain. Here we present the functional assignment and biochemical characterization of seven proteins, Cpr33-38 and Cpr12, that initiate the biosynthesis of the polyamide.

Functional characterization of Cpr38, which has sequence similarity to the gene products encoded by *pabA* and *pabB* from *E. coli*, revealed that it functions as a 4-amino-4-deoxychorismate (ADC) synthase catalyzing a two-step reaction involving amidohydrolysis of L-Gln with ammonia channeled and incorporated into chorismic acid to generate ADC. Cpr12, encoded by a gene that was originally proposed to be outside the gene cluster and sharing similarity to proteins annotated as ADC lyase, was revealed to catalyze the elimination of pyruvate to form PABA.

Cpr36 is demonstrated to function as a free-standingpeptidyl carrier protein (PCP), which is activated to form holo-protein from the apo-form. Cpr37, which belongs to the adenylation domain protein in the nonribosomal peptide synthase (NRPS), subsequently activates PABA and loads it to holo-Cpr36 Two proteins Cpr34 and Cpr35 work in concert to catalyze decarboxylative condensation between a thioester linked PABA and malonyl-*S*-acyl carrier protein (ACP) during aromatic polyketide biosynthesis catalyzed by type II polyketide synthases.

Following condensation, Cpr33 acts as 3-oxoacyl-ACP reductase that catalyzes reduction to the β-hydroxythioester intermediate. In this scenario, hydride is predicted to be added to the *re* face to generate the *S* configuration resulting in the same

stereochemical outcome as other 3-oxoacyl-ACP reductase (FabG) from bacterial type II fatty acid synthases.These findings are critical advancement for interrogating the biosynthesis of the unusual chemical components of the family of antibiotics of capuramycin.

KEWWORDS: capuramycins, peptidoglycan cell wall, polyamide assembly

Erfu Yan
Student's Signature
08/28/2018
Date

BIOSYNTHETIC MECHANISM OF THE ANTIBIOTIC CAPURAMYCIN

By

Erfu Yan

Steven Van Lanen
Director of Dissertation

David Feola
Director of Graduate Studies

08/28/2018
Date

献给我的妻子和女儿，婷婷和雯潇，谢谢你们一路相濡以沫

This dissertation is dedicated to my dear wife and daughter, Tingting and Liliana, whose love and support made this journey through graduate school possible.

ACKNOWLEDGMENTS

I would like to take this special moment to express my gratitude to my mentor, Dr. Steven Van Lanen, for his guidance, support, and inspiration throughout my study in his group. I am grateful for the chance to seek refuge in this lab following departure of Dr. Guo's lab in my second PhD year. His expertise and insightful discussions have benefited me greatly and extensively. I would also like to thank my committee members, Dr. Chang-guo Zhan, Dr. Oleg Tsodikov, and Dr. Chris Richards, for their invaluable advice and comments over the years, and I would like to sincerely thank Dr. Zhenyu Li for being the external department examiner for my dissertation defense.

Special thanks to all current and previous members from the Van Lanen group: Dr., Dr. Zhaoyong Yang, Dr. Sandra Barnard, Dr. Xiuling Chi, Dr. Anwesha Goswami, Dr. Xiaodong Liu, Dr. Zheng Cui, Dr. Ying Huang, Matthew McErlean, Ashley Arlinghaus, and Jonathon Overbay. Also, a special thanks to Dr. Bert C. Lynn and Dr. Fang Huang in Department of Chemistry, University of Kentucky. I would also like to thank Dr. Minakshi Bhardwaj for analyzing all the NMR data showed in this thesis. It was a great honor to work with and learn from you all!

Last but not least, I would like to acknowledge my family and friends for their constant support and encouragement. I need to thank all my friends and coworkers who have been my teachers in different aspects of my life in graduate school.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

7-AAD 7-aminoactinomycin

ACL aminocaprolactam

ACP acyl carrier protein

ADC aminodeoxychorismate

ADCS 4-amino-4-deoxychorismate synthase

ADCL 4-amino-4-deoxychorismate lyase

ADME absorption, distribution, metabolism and excretion

Ala Alanine

amp ampicillin

apr apramycin

ATP adenosine-5′-triphosphate

AS anthranilate synthase

Asp Aspartic acid

BLAST basic local alignment search tool

CarU uridine-5'-carboxamide

CDC Center for Disease Control and Prevention

chl chloramphenicol

DAP diaminopimelic acid

DCM dichloromethane

Ddl D-Ala-D-Ala ligase

DIG digoxigenin

DMF dimethylformamide

DMSO dimethylsulfoxide

DNA deoxyribonucleic acid

E. coli Escherichia coli

EC_{50} half maximal effective concentration

EDTA ethylenediaminetetraacetic acid

FRT flippase recognition target

FDA US Food and Drug Administration

4′-PPT 4′-phosphopantetheine

GDP guanosine-5′-diphosphate

GlcNAc *N*-acetylglucosamine

GlyU (5′*S*,6′*S*)-5'-*C*-glycyluridine

Glu Glutamic acid

GTP guanosine-5′-triphosphate

His6 hexahistidine

HMBC heteronuclear multiple-bond correlation spectroscopy

HPLC high performance liquid chromatography

h hour

HRMS high-resolution mass spectrometry

HSQC heteronuclear single-quantum correlation spectroscopy

HTS high-throughput screening

IC_{50} half maximal inhibitory concentration

IPTG isopropyl-β-D-thiogalactopyranoside

ICS isochorismate synthase

k kilo

kan kanamycin

kb kilo base pairs

k_{cat} turnover rate

kDa kilo Dalton

K_d dissociation constant

K_M Michaelis-Menten constant

LB Luria broth

LC-MS liquid chromatography–mass spectrometry

Lys lysine

MBP Maltose binding protein

MDR multiple drug resistant

MIC minimum inhibitory concentration

min minute

MraY phospho-MurNAc-pentapeptide translocase

MRSA methicillin-resistant *Staphylococcus aureus*

MurNAc *N*-acetylmuramic acid

MW molecular weight

NADP+ nicotinamide adenine dinucleotide phosphate

NADPH reduced form of NADP+

NDP nucleoside diphosphate

NME new molecular entities

NMR nuclear magnetic resonance spectroscopy

NRP nonribosomal peptides

NRPS nonribosomal peptide synthase

NTP nucleoside-5′-triphosphate

OD optical density

ORF open reading frame

OriT origin of transfer

PABA para-aminobenzoic acid

PAGE polyacrylamide gel electrophoresis

PBP penicillin binding protein

PCP peptidyl carrier protein

PCR polymerase chain reaction

PEP phosphoenolpyruvic acid or phosphoenolpyruvate

pI isoelectric point

PLP pyridoxal-5'-phosphate

RNA ribonucleic acid

PNPT polyprenyl-phosphate N-acetyl hexosamine 1-phosphate transferase

rt room temperature

SAR structure-activity relationship

SDR short chain dehydrogenase/reductase

SDS sodium dodecyl sulfate

SDS-PAGE sodium dodecyl sulfate polyacrylamide gel electrophoresis

SHMT serine hydroxymethyltransferase

Svp *Streptomyces verticillus* phosphopantetheinyltransferase

sp. Species

TCA Trichloroacetic acid

TB tuberculosis

TEA triethylamine

TFA trifluoroacetic acid

THF tetrahydrofolic acid

UA uridine-5'-aldehyde

UDP uridine-5′-diphosphate

UDP-GlcNAc uridine-5′-diphospho-N-acetylglucosamine

UDP-MurNAc uridine-5′-diphospho- N-acetylmuramic acid

UMP uridine-5′-monophosphate

UTP uridine-5′-triphosphate

UV/Vis ultraviolet/visible

VRE vancomycin-resistant Enterococci

WT wildtype

XDR TB extensively drug-resistant tuberculosis

α-KG α-ketoglutarate

Chapter 1: Introduction and Background

1.1 Natural Products – Significance

The exploration of natural products and their derivatives has traditionally been a successful strategy in the discovery of novel medicines and therapeutics [1]. Before the advent of high-throughput screening and the post-genomic, more than 80% of drug substances were natural products or inspired by a natural compound [2]. In the last ~20 years, almost half of the drugs approved by FDA are based on natural products [3]. The main reasons why natural products have always been important for the pharmaceutical industry are included but not limited to: a) most of the classical drugs either contain natural products or have these as original leads, i.e., there are no known synthetic substitutes to continue to inspire synthetic and analytical chemists for nature products [4], b) natural products continue to encourage organic chemists typically make compounds by controlling the reactivity of far more diversified 'functional groups', remaining a significant source of human medicines [5], c) natural products are invaluable tools for deciphering complex metabolic pathways and leading to critical biomedical insights [6], and d) there is still unlimited potentiality to discover novel therapeutics from unexplored natural products sources [7, 8].

Although the discovery of new natural products has declined the past two decades as large pharma have shifted their research efforts from natural products discovery to the more profitable drug candidates, the use of natural products as sources of therapeutics still play a highly predominant role in medicine. In general, almost half of new drug molecules discovered and developed in last 30 years were inspired or originate from natural products [9]. In the meantime, the other half of new molecular entities were from synthetic origin. As for the drugs that kill different kinds of bacteria, approximately 75% of approved antibacterial compounds were derived from natural products [9]. Natural products play an even more critical role to fight against human diseases. As a result, there are emerging trends that research efforts in the field of

natural products exploring have significantly increased in academia and specialized pharmaceutical or biotechnological companies with improved technologies of screening and developing to front-line drug candidates.

In most case, natural products often need to be modified to improve clinical properties and bypass resistance mechanisms, which increase the structure complexity of natural products [10]. The approach of semisynthesis was used to create these products and improve upon the clinical utility [11]. Apart from semisynthesis, another synthetic strategy of protein engineering was also used to accelerate the discovery of a wide range of structural diversity. To develop more technologies which can be used to modify the genetic templates that specify these natural products which can be used for human disease treatment, the genetic and biochemical studies to define and discovery of novel biosynthesis and manipulation mechanism of natural products synthesis are highly in demand.

1.2 Need for New Anti-tuberculosis Antibiotics

Although there has been the remarkable success of antibiotics, infectious diseases are still one of the threats to human health worldwide nowadays. This problem is mainly due to that bacteria, and other pathogens such as fungi, protists and other non-bacterial pathogens have always evolved so that they can resist new drugs as a result of both intrinsic or acquired mechanisms. Overall, it is not just a problem confined to bacteria, but all microbes that have the potential to render drugs ineffective.

Drug discovery and development efforts of new antibiotics are not keeping pace while antibiotic use is rising [12]. In 1950s, most classes of antibiotics were discovered and it has been described as the “golden era” of antibiotics. Since then, the discovery and development of new antibiotics declined radically. Nonetheless, natural products or derivatives of natural products remain one of the major resources for human drug discovery and contributed about 75% of the total FDA-approved antibacterial agents. Between 1983 and 1987, there were 16 new systemic antibacterial agents were

approved by the US Food and Drug Administration (FDA), but this number has decreased dramatically to 2 between 2008 and 2012 [12].

Bacterial resistance to every major class of antibiotic has led to the emergence of multiple drug-resistant pathogens (MDR) such as methicillin-resistant *Staphylococcus aureus* (MRSA), and vancomycin-resistant *Enterococci* (VRE) has become severe hospital and community threat. What's more, it was determined that 700 000 deaths worldwide are attributable to antimicrobial resistance (AMR). Predictions show that 10 million people will die worldwide each year by 2050 due to increasing AMR, with the highest potential burdens in Asia and Africa [12]. Therefore, discovery and development of new antibiotics which can kill such kind of drug-resistant pathogens remains a top priority.

Tuberculosis (TB) is a disease caused primarily by *Mycobacterium tuberculosis*. The cell wall of *M. tuberculosis* is unique compared to other bacteria and is composed of upper and lower segments. The upper segment is composed of free lipids and some with shorter fatty acids complementing the longer chains. The lower segment cell wall core forms the mycolyl-arabinogalactan–peptidoglycan complex. This complex structure makes it naturally resistant to most antibiotics. Specifically, the cell-wall architecture is composed of peptidoglycan and several strands of highly branched arabinofuran which is further attached to mycolic acids [13]. The thick layer of lipid located on the outer part of the cell protects the *tuberculosis bacillus* from deleterious chemicals and allows the bacterium to survive by living inside macrophages [14]. All of these unique and intricate structures make TB difficult to treat [15]. TB is estimated to currently infect 2 billion people throughout the world [16]. An estimated 1.3 million deaths are associated with *Mycobacterium tuberculosis* infection annually [17], making it the second largest infectious disease killer all over the world. According to Centers for Disease Control and Prevention's 2013 report, there were 8.6 million new TB cases globally with a reported 1.3 million deaths in the year of 2012 [15]. While in the United States, the diseases like multidrug-resistant and extensively drug-resistant *tuberculosis*

(MDR and XDR TB) are an increasing threat. In the year of 2012, an estimated 4,500 people were reported to have acquired MDR-TB in the United States [17]. Identical to other pathogens, there has been a steady increase in reported cases of drug-resistant *M. tuberculosis* each year [17]. All of this makes TB a significant threat to global health. Therefore, the discovery and development of new anti-tuberculosis antibiotics are urgently needed.

Since the rise of antibiotic resistance all over the world, it is critical to keep discovering antibiotics and many strategies has been used to achieve this goal. Whole cell-based high-throughput screening (HTS) is one technology which has resulted in the discovery of a large number of bioactive compounds. However, it remains difficult to determine the targets of novel compounds, and the bioactive fractions can often be toxic to mammalian cells [18]. Traditional whole cell-based HTS has almost reached the point of diminishing returns. For this reason, a target-specific HTS model for *in vitro* screening offers a new path for the discovery of novel antibiotic candidates as well as specific binding targets. Traditionally, to improve the pharmacological properties such as absorption, distribution, metabolism and excretion (ADME) of drug candidates, researchers need to compromise their target-specificity. However, the resulting hit-target pairs in HTS [18]can be further advanced by structure-based drug design. Excluding target-specificity, HTS model can address the bottleneck of species-specificity. Species-specific antibiotics are significant as they typically minimize the adverse effects of human therapy. It is quite normal that even subtle differences in targets among different species may lead to false-positive screening results. For example, there were studies revealed that bedaquiline selectively inhibits the growth of mycobacteria but shows no activity against other bacteria owing to minor differences in the target ATP synthase of mycobacteria and other species [19-22].

1.3 Biosynthesis of Peptidoglycan Cell Wall

Peptidoglycan is the primary constituent of bacterial cell walls and is crucial for the

survival of both gram-positive and gram-negative bacteria [23]. In the 1940s, bacterial peptidoglycan and its precursors were identified and characterized [24]. The biosynthesis pathway of peptidoglycan is one of the well studied targets for the design and development of antibiotics [23, 25, 26]. The structure of peptidoglycan consists Nacetylmuramic acid (MurNAc) and N-acetylglucosamine (GlcNAc) linked by β-1,4-linked polysaccharide peptidyl bonds (Figure 1.1).

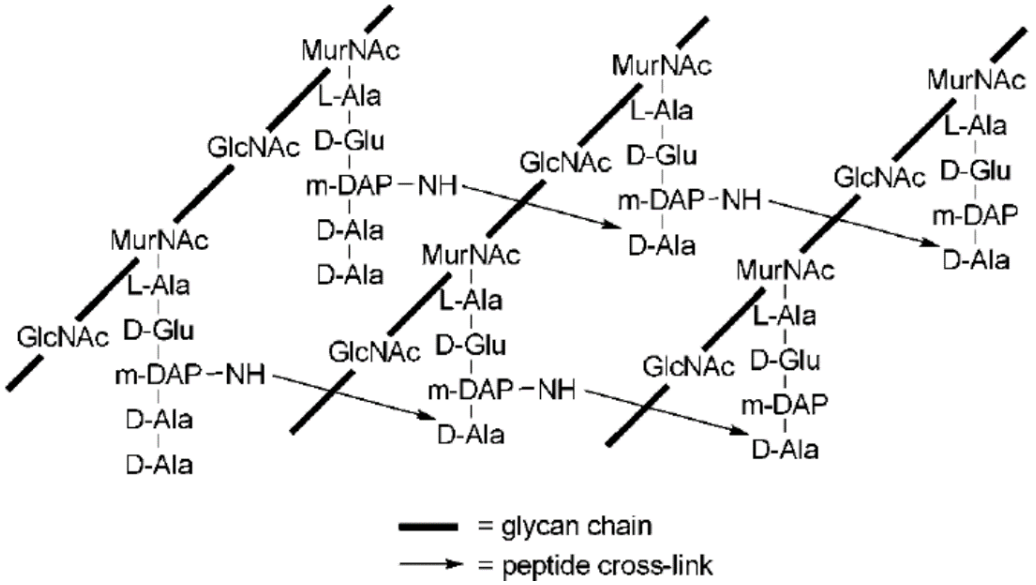

Figure 1.1 Structure of peptidoglycan crosslinking [27].

A short polypeptide of five amino acids of L-Ala-γ-D-Glu-X-D-Ala-D-Ala where X is either L-Lys or meso-diaminopimelic acid (DAP) is attached to the 3-position of the MurNAc sugar chain. This unique structure leads to cross-linking between the 3-amino acid and the peptide bond of another polymeric unit. These crosslinks provide the structural integrity that allows the cell wall to withstand the osmotic pressure of the cytoplasm. Overall, the biosynthesis of peptidoglycan requires a minimum of 12 ubiquitous enzymatic reactions in both Gram-positive and Gram-negative bacteria [25, 27]. (Figure 1.2).

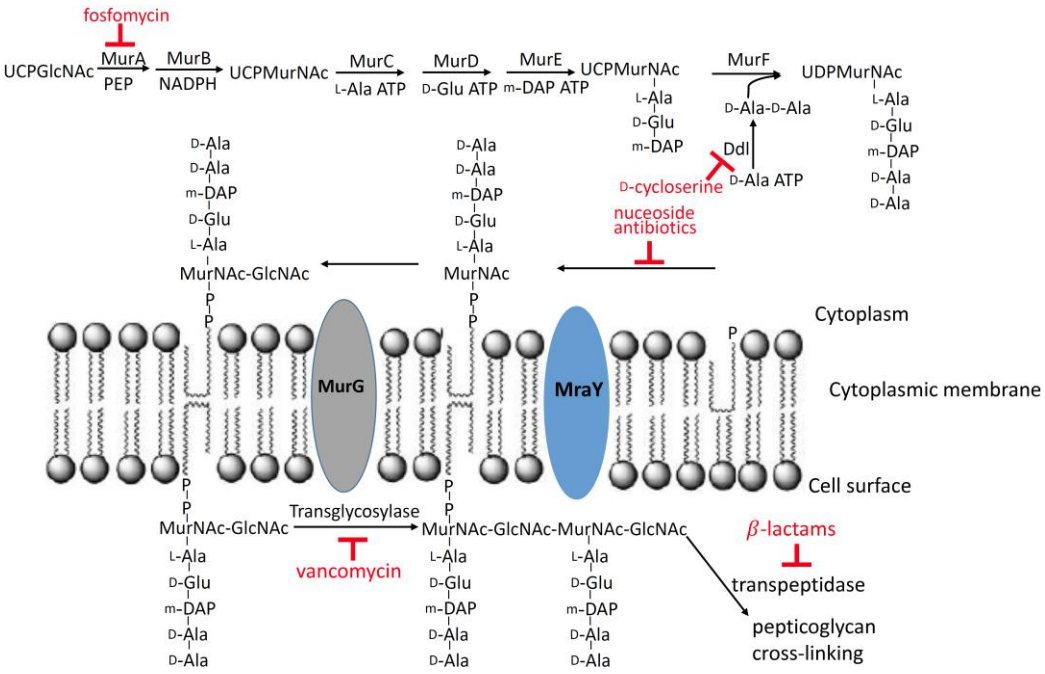

Figure 1.2 Summary of peptidoglycan biosynthesis. Antibiotics that inhibit specific enzymes are highlighted in red (vancomycin inhibits transglycosylase and β-lactams inhibits transpeptidase by binding to the terminal units of peptidoglycan).

In the initial cytosolic stage, uridine-5′-phosphate (UDP)-sugars are the biosynthetic precursors for the biosynthesis of all cell wall components. Initially, MurA and MurB catalyze the transformation of uridine-5-diphospho-N-acetylglucosamine (UDP-GlcNAc) to UDP-N-acetylmuramic acid (UDP- MurNAc), followed by attachment of phosphoenolpyruvate (PEP). A series of ATP-dependent ligases MurC–F catalyze the addition of polypeptide chain sequentially to UDP-MurNAc to generate UDP-MurNAc-pentapeptide, which is the final cytoplasmic precursor (Figure 1.2) [23, 25, 26, 28].

The UDP-MurNAc-pentapeptide provides phospho-MurNAc-pentapeptide, which is transferred to undecaprenyl phosphate by the enzyme translocase I (phosphor-MurNAc-pentapeptide translocase) or MraY. Addition of GlcNAc sugar onto the 4′-hydroxyl group of MurNAc lipid intermediate I, catalyzed by the glycotransferase MurG, leads the formation of lipid intermediate II [29]. In certain gram-positive strains, a unique aminoacyl-tRNA donor is utilized to incorporate amino acids to lipid

II (for examples, five Gly residues are added on in *Staphylococcus aureus* [30, 31]. The lipid intermediate is flipped from cytosolic side to outside of the cell membrane by a membrane protein called 'flippase' [31]. On the cell surface, Lipid II is polymerized by transglycosylation and transpeptidation, the latter of which is catalyzed by penicillin binding proteins (PBPs) [32]. Undecaprenyl pyrophosphate is released during the polymerization and recycled via enzymatic dephosphorylation [31]. Since most of the enzymes that participate in the biosynthesis of peptidoglycan lack mammalian homologs, cell wall biosynthesis has been proven targets for inhibiting bacterial survival and growth [23]. There have been some clinically tested natural products that inhibit the specific steps (highlighted in red in figure 1.2). Within the cytosolic steps, natural products fosfomycin and D-cycloserine targeted the enzyme of MurA and DdB [23], while MurG is inhibited by peptides ramoplanin and enduracin [33]. Other than the inhibitors for the enzymes in cytosolic steps, there are also some target inhibitors. Bacitracin targets the lipid carrier itself by binding irreversibly to undecaprenyl pyrophosphate [30]. Vancomycin inhibits the transglycolsylase by binding to the substrate itself [34], and the β-lactams irreversibly inhibit transpeptidase activity [32]. Unlike most bacteria, the mycobacterial cell wall peptidoglycan has exceptional features including densely cross-linked DAP-DAP bridge and incorporation of N-glycolyl-muramic acid in polysaccharide chains [35]. Although these inhibitors have been successfully developed as antibacterial antibiotics, many of them have issues of resistance within years of clinical usage. Notably, among these six families of inhibitors, none of them have been clinically proven effective against TB. Only the natural products of bacterial translocase I inhibitor, haven't been used in clinic yet, making it an ideal target for new antibiotic discovery and development [32].

1.4 MraY – Structure and Function

Phospho-MurNAc-pentapeptide translocase I (MraY) is an integral membrane enzyme that is essential for peptidoglycan biosynthesis. The catalytic reaction of MraY is the

transfer of phospho-MurNAc-pentapeptide from UDP-MurNAc-pentapeptide to the lipid carrier undecaprenyl-pyrophosphate, releasing undecaprenyl-pyrophosphoryl-MurNAc-pentapeptide (also known as Lipid I) (Figure 1.3) [29, 36].

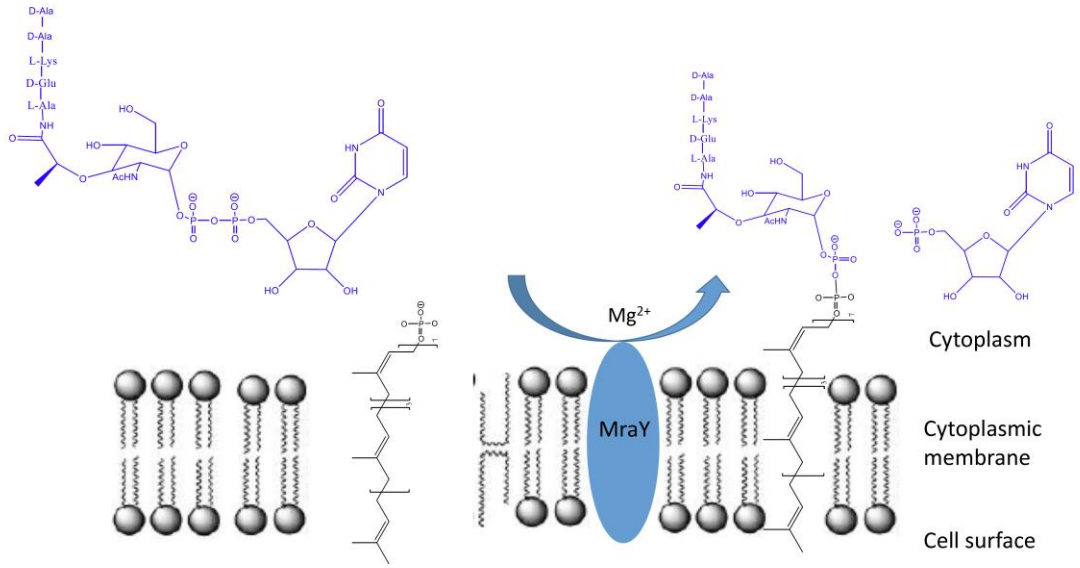

Figure 1.3 Reaction catalyzed by MraY. MraY is a transmembrane protein responsible for initiating of peptidoglycan biosynthesis.

This catalytic step is Mg^{2+}-dependent and essential for all bacterial viability, and therefore represents a novel target for the development of new classes of antibiotics [29]. The transferase activity of MraY was first discovered by Neuhaus in 1965 [37]. However, the gene for MraY was not identified until 1991 when it was over-expressed in *E. coli*, the gene product was demonstrated to have a significantly enhanced transferase activity relative to the wild-type strain [36].

MraY belongs to a subfamily of the polyprenyl-phosphate N-acetyl hexosamine 1-phosphate transferase (PNPT) superfamily. This superfamily includes enzyme responsible for the synthesis of cell envelope polymers such as the O-antigen and teichoic acid in bacteria, and the GPT (UDP-GlcNAc:dolichol-P GlcNAc-1-P transferase) enzyme family responsible for catalyzing the transfer of GlcNAc-1-P from UDP-GlcNAc to dolichol-P to form GlcNAc-P-P-dolichol in eukaryotes [35, 38].

Bioinformatics analysis of MraY by BLAST shows that it has sequence similarity to

UDP-sugars transferase enzymes from both eukaryotes and prokaryotes [39, 40]. Studies conducted with MraY isolated from *Staphylococcus aureus* and *Micrococcus luteus* [41] in addtion to MraY from *E. coli* has supported that the enzyme acts as a transmembrane protein vital for cell survival and Mg^{2+} is required for activity [28, 42]. Two catalytic mechanisms for MraY have been proposed. One mechanism involves a single-step displacement while one invoked a two-step, glycosylated enzyme intermediate. Isotope exchange experiments conducted by Neuhaus et al. suggested that that the MraY-catalysed reaction is a two-step reaction, wherein S_N2 type nucleophilic substitution forms a covalent enzyme intermediate within the reaction [37].

Subequent biochemical characterization of MraY has been hindered, however, due the numerous technical difficulties associated with the preparation of both substrates and the enzyme. For example, MraY from *E. coli* has been overexpressed, extracted, and characterized kinetically using a continuous fluorescence assay, but it could not be purified to homogeneity [40]. MraY from *B. subtili*s has also been purified to apparent homogeneity in small quantities, although the specific activity is significantly decreased in these preps. A method of purification of recombinant *E. coli* MraY has been developed by using an engineered C-terminus His6-tag, which lead to a more active protein [43]. Furthermore, Muramatsu et al. have isolated MraY enzyme from *Bacillus subtilis* and shown it to be active, however, similar to *E. coli* MraY, the specific activity was lower compared to crude preparations [44].

The *mraY* gene was shown to be essential for viability in gram-positive bacteria *Streptococcus pneumonia*. The bioinformatics analysis of 879 sequenced microbial genomes have been conducted and suggest that only a single copy of the essential *mraY* gene is encoded per genome [31]. Studies conducted in *E. coli* involving mutational inactivation of the *mraY* gene led to growth inhibition and lethal phenotype with early growth characteristics which typically observed from similar strains containing inactivated genes involved in peptidoglycan biosynthesis [39].

The structural basis for the enzyme's function was obscure until recently. It was initially

predicted that MraY is a transmembrane protein with ten α-helical regions, six periplasmic segments and five cytoplasmic segments. Ten α-helices are expected to form the hydrophobic tunnel that would correspond to the binding domain of undecaprenol phosphate. Five cytoplasmic segments are arranged to bear essential amino acids required for substrate recognition and catalysis [27] (Figure 1.4A). Recent mutational studies conducted by Thomas et al. established 14 partial phage resistance phenotype associated invariant charged amino acid residues which are essential for enzyme activity in *B. subtilis* MraY [45], most of these residues are found in a specific 'cleft' region which was predicted to function as the active site.

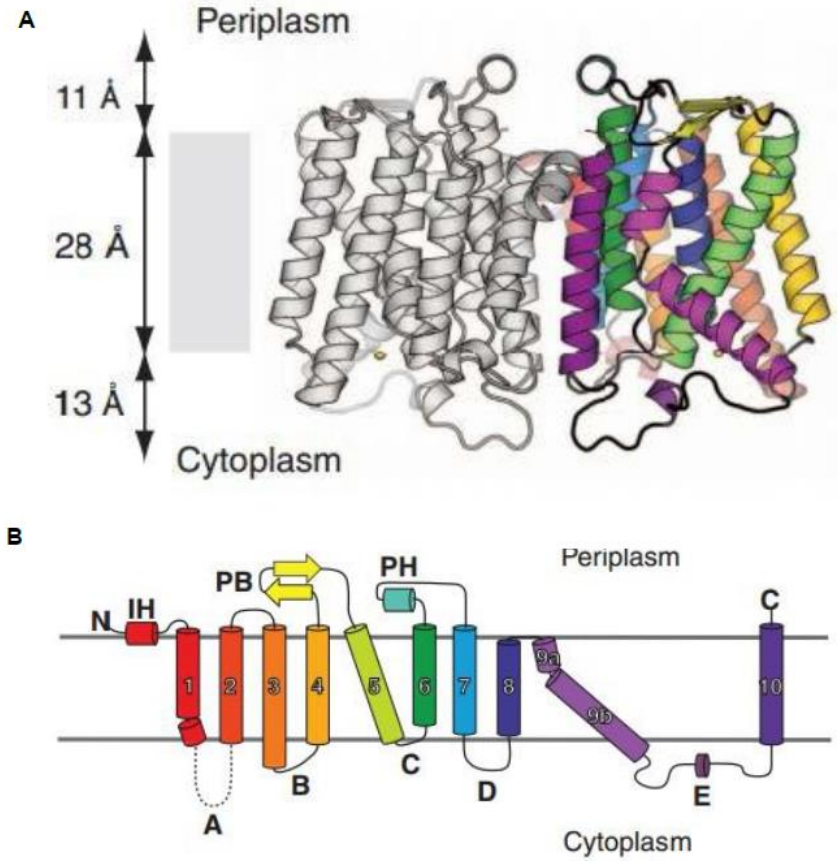

Figure 1.4 Structure of MraYAA[40].

Bouhss et al. identified topological MraY structure by taking advantage of β-lactamase mraY-blaM fusion system with both enzymes purified from *S. aureus* and *E.coli* [42]. Recently Lee et al. report the crystal structure of MraY from *Aquifex aeolicus*

(MraYAA), the first structure of the PNPT superfamily, at a 3.3Å resolution [40](Figure 1.4A). MraYAA has been shown to crystallize as a dimer and at the center of the dimer interface is an oval-shaped tunnel. The tunnel is surrounded mostly by the hydrophobic amino acids and is large enough to presumably accommodate lipids [40].

To study the specific residue function, Asp-115, Asp-116 and Asp-267 of MraYAA were examined using site-directed mutagenesis. These three aspartic acid residues are conserved within all MraYs and are found on the cytoplasmic side of the lipid bilayer. Asp-267 has been proposed to be the active site residue that is used in the formation of a covalent enzyme-phospho-MurNAc-pentapeptide intermediate. Asp-115 and Asp-116 may form a binding site for Mg^{2+} (Fig 1.4b) [27, 43]. Identical results were obtained upon mutational inactivation of three conserved aspartate residues along with a conserved histidine residue of MraYAA (Asp117, Asp118, Asp265, and His324), each of which resulted in the complete loss of activity. For MraYAA, however, Asp265 was predicted to be involved in Mg^{2+} binding while Asp117 was regarded as a specific base that deprotonates the undecaprenyl phosphate (Figure 1.5).

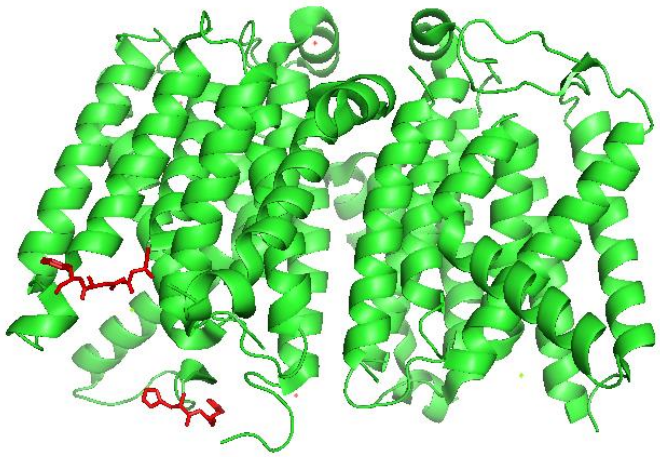

Figure 1.5 The active site of MraYAA [40]. Amino acids important for catalysis are colored in red (figure was redrawn using published data, PDB ID: 4J72).

1.5 Inhibitors of MraY-Translocase I

MraY is regarded as a promising target for the development of new antibiotics because the Mg^{2+} dependent reaction is essential for bacterial viability of both gram-negative and gram-positive bacteria [38]. In the past several decades, several MraY inhibitors have been discovered, including several different classes of natural product inhibitors with antibacterial activity [45, 46]. There are three distinct types of compounds that can inhibit MraY activity, namely protein E, lipopeptides and nucleoside antibiotics (Figure 1.6).

Protein E is a helical polypeptide composed of 91 amino acids encoded by DNA phage φ-174 that leads to bacteriolysis. Recently, Tanaka et al. have shown that 18-residue peptides mimicking the wild-type sequence of the conserved transmembrane helix of protein E are sufficient to lyse host cells, suggesting that this helix is essential for activity [47]. In a different study, Sharon et al. demonstrated that a 37-amino acid polypeptide comprising the transmembrane domain of protein E was sufficient to

specifically bind and inhibit MraY [48].

The second category of MraY inhibitors is the lipopeptides including amphomycin, friulimcin and glycinocins [29]. These compounds inhibit MraY activity by complexing with undecaprenyl phosphate in the presence of Ca^{2+}, a mechanism analogous to the mode of interaction of glycopeptides like daptomycin [49, 50]. Amphomycin has excellent activity against Gram-positive bacteria enterococci and streptococci [51, 52].

The last category of MraY inhibitors is nucleoside antibiotics, which are the largest and most structurally diverse group of inhibitors. As a result of structural variations, these nucleoside inhibitors are further classified into four structural families: 1)the peptidyl nucleosides represented by pacidamycin from *Streptomyces coeruleorubidus* [53] and mureidomycin from *Streptomyces flavidovirens* [54], 2) the lipodisaccharyl nucleosides represented by tunicamycins from *Streptomyces lysosuperificus* [55], 3) the glycosyl-peptidyl nucleosides represented by the capuramycins A-503083s from *Streptomyces* sp. SANK 62799 [56], A-500359s from *Streptomyces griseus* SANK 60196 [57], and A-102395 from *Amycolatopsis sp.* SANK 60206 [58], and 4) the lipopeptidyl nucleosides, which includes A-90289s from *Streptomyces sp.* SANK 60405 [59], FR-900493 from *Bacillus cereus* No. 2045 [60, 61], caprazamycins from *Streptomyces sp.* MK739-62F [62], and muraymycins from *Streptomyces sp.* NRRL 30471 [60]. All nucleoside antibiotics contain a modified uridine nucleoside that has been shown to be critical for their biological activities. This nucleoside component can consist of a hexofuranoside (C6), a heptafuranoside (C7), a peptidyl moiety (C5) or an undecafuranoside (C11). The structural variability leads to differences in the specific pharmacology mechanism against MraY [54, 63]. Peptidyl nucleosides such as mureidomycin, has been shown to be a competitive inhibitor with respect to substrate UDP-MurNAc-pentapeptide. Unfortunately, tunicamycins are also established as inhibitors of the UDP-GlcNAc:dolichol-P GlcNAc-1-P transferase (GPT) enzyme family which plays a crucial role in glycosylation in eukaryotes. Therefore tunicamycins display toxicity, making them unattractive leads for antibiotic

development [64, 65].

On the other hand, the lipopeptidyl nucleoside family of MraY inhibitors, which includes A-90289s from *Streptomyces sp.* SANK 60405 [59], FR-900493 from *Bacillus cereus* No. 2045 [60, 61], caprazamycins from *Streptomyces sp.* MK739-62F [62], and muraymycins from *Streptomyces sp.* NRRL 30471 [60], contains an aminoribosyl moiety—a 5-amino-5-deoxyribose—which is attached to a heptofuranose nucleoside component called 5′-C-glycyluridine. Structure-activity relationship studies using simplified synthetic analogues have shown that the glycyluridyl and aminoribosyl moieties are both critical for the antibiotic activity [66, 67].

Glycosyl peptidyl nucleosides, represented by the capuramycins, show a mixed-type inhibition against UDP-N-MurNAc pentapeptide [68]whereas peptidyl, lipodisaccharyl and lipopeptidyl nucleosides are competitive MraY inhibitors [69]. Even though variability associated with the preparation and assay of MraY may lead to the variations in mechanism of inhibition, it is certain that the nucleoside antibiotics described herein selectively inhibit the MraY-catalyzed reaction [28]. The structure of capuramycin contains three distinct moieties: uridine-5'-carboxamide (CarU), a rare unsaturated hexuronic acid, and an aminocaprolactam (LACL). Comparing to other two members of capuramycins A-500358s and A-500359s, A-102395, however lack the aminoribose but contains a unique caprolactam moiety which has been shown to be critical to its activity [57, 68].

The biosynthesis of capuramycins A-500358s and A-500359s have been studied very well. On the other hand, little is known about the biosynthesis of A-102395, which is the primary topic of this thesis (Figure 1.6).

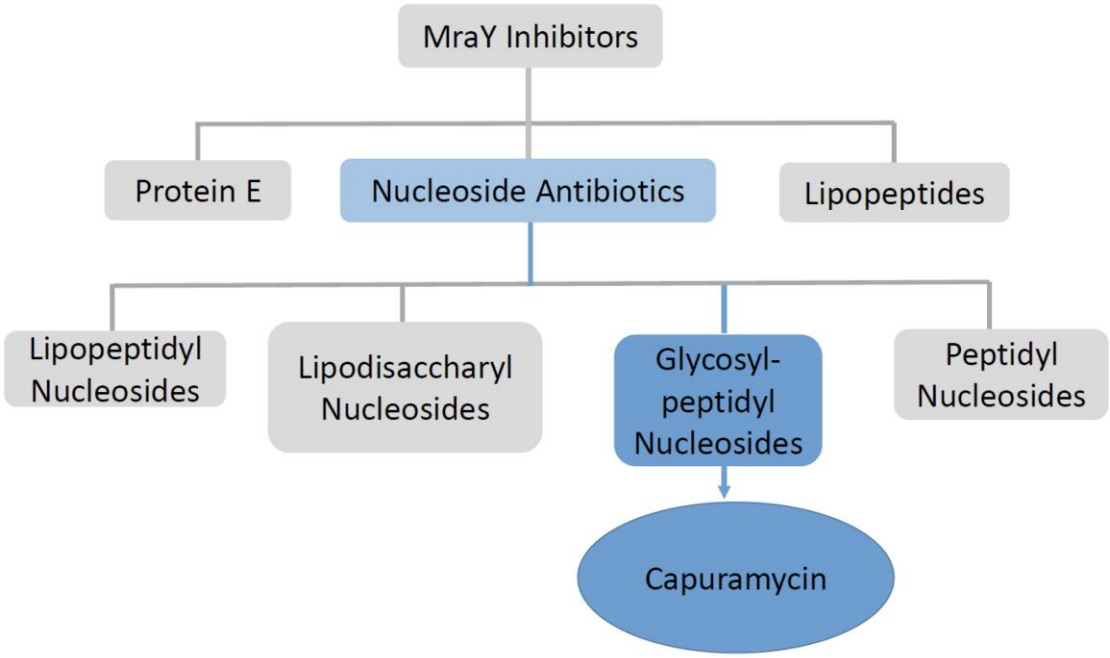

Figure 1.6 Representatives of different classes of MraY inhibitors.

1.6 Non-ribosomal Peptides and Their Synthesis

Non-ribosomal peptides (NRP) are a large pool of biologically active natural small peptides synthesized mainly by bacteria and fungi. NRP have a wide range of clinic applications and are involved in the production of some of our most crucial antibacterial, antifungal, antiviral, immunosuppressant, and anticancer drugs. Non-ribosomal peptides exhibit high structural diversity with almost one third being linear; the remainder having cyclic, branched or other complex primary structures [70]. The diverse structures can explain this diverse bioactivity regarding the monomers that can be incorporated [70].

Significant progress has been made in the past six decades toward understanding the synthesis of non-ribosomal peptides. In the 1950s, researchers were beginning to decipher the mechanism of ribosome-dependent protein synthesis and found that a distinct enzymatic mechanism from the ribosome involved in the production of NRP.

Tatum et al. provided the first evidence that tyrocidine was biosynthesized by a mechanism independent of the ribosome [71-73]. The concept of a modular enzymatic mechanism for non-ribosomal peptides synthases (NRPSs) was valuable in interpreting genetic information once DNA sequencing became routine. Later on, Marahiel et al. led the effort to sequence the biosynthetic gene clusters coding for the NRPSs that generate the unusual peptides from *Bacillus* species that initiated the analysis of this type of enzymology. In 1994, Stein proposed that NRP synthesis follows the multiple template model. According to this model, a line-like manner was used to synthesis the peptides in a modular assembly by NRPS enzymes. The modules are located within multiple distinct enzymes that associate post-translation and classified as either initiation, elongation, or termination modules depending on their location in the assembly line [74](Figure 1.7). Acting in a concerted but semiautonomous fashion, modules are defined by their ability to recognize, activate and incorporate a specific monomer into the final peptide product [75].

Specifically, the assembly of a nonribosomal peptide involves a series of repeating steps that are catalyzed by three core catalytic domains: the adenylation, thiolation (or peptidyl carrier protein), and condensation domains. An adenylation (A) domain recognizes and activates a specific substrate by addition of AMP within each module (Figure 1.7A). Once the aminoacyl-AMP intermediate is formed, the A domain conjugated with a partner T (or peptidyl carrier protein, PCP) domain to create the aminoacylthioester intermediate. The activated substrate is then tethered to a flexible 4′-phosphopantetheine (4′-PPT) prosthetic group, which is itself covalently attached to a thiolation (T) domain (also known as a peptidyl carrier protein (PCP) domain) (Figure 1.7B). With its flexible PPT prosthesis efficiently the ‘swinging arm’ of a biomolecular assembly line, the T domain transfers peptide intermediates between different domains and modules. Post-attachment of an activated substrate by its adenylation (A) domain partner, a thiolation(T) domain then passes that substrate to a condensation (C) domain. A peptide bond was formed between the donor substrate provided by the T

domain immediately upstream and the acceptor substrate provided by the downstream T domain (Figure 1.7C). T domains belong to the same superfamily of proteins as acyl carrier proteins (ACP) involved in polyketide and fatty acid biosynthesis, and they all share the same four-helix bundle structure [76]. Like ACPs, T domains must be post-translationally modified with a 4′-PPT arm. 4′-PPT transferase activate T domain by the addition of the 4′-PPT on a conserved serine residue [77]. In the 1960s, Lipmann et al. noted that the terminal thiol of the 4′-PPT group attacks the carboxyl group of the aminoacyl-AMP intermediate resulting in the formation of the aminoacylthioester. Dario et al. applied a unique top-down based tandem mass spectrometry (MS) approach to study the phophopantetheinylation of the acyl carrier protein after incubation with CoA in the presence or absence of Sfp [78].

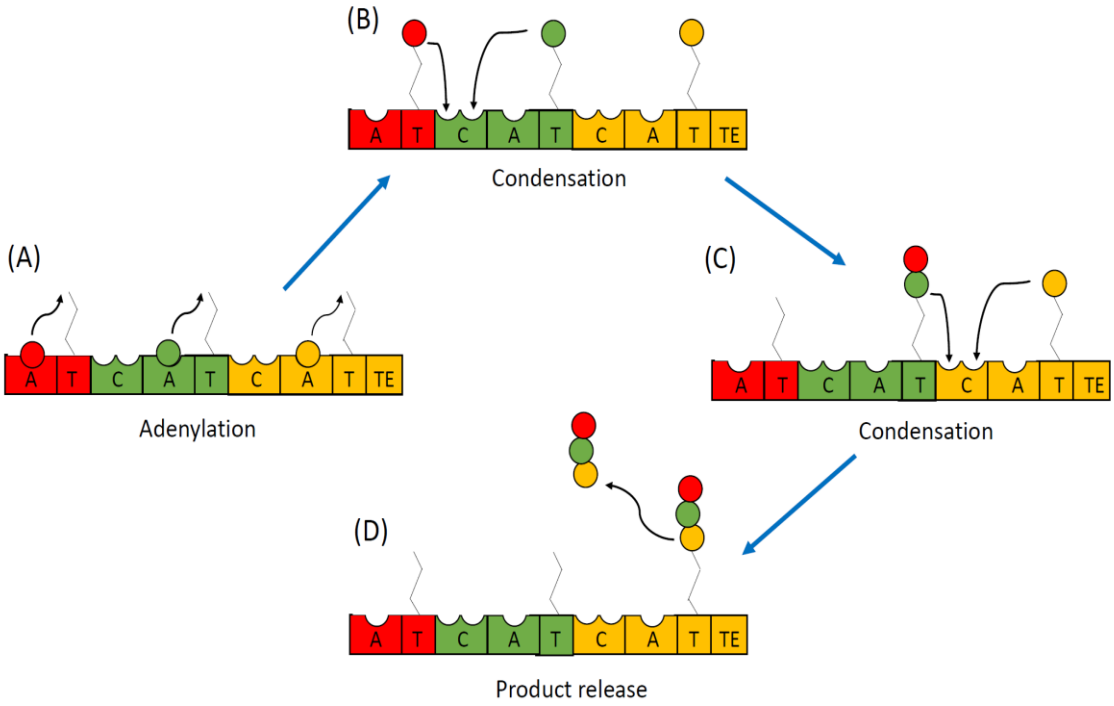

Figure 1.7 Simplified NRPS modular mode of peptide synthesis.

Following the initial condensation event, the process can repeat iteratively, with the previous peptide intermediate now serving as the donor substrate for the C domain.

In this pathway, some additional tailoring domains modify individual substrates in a directed fashion (e.g., epimerization, (E) domains, for conversion from L- to D-

enantiomers). The growing peptide continues to be passed from the T domain of one module to the T domain of the next until the product is released (Figure 1.7D) [75, 79].

1.7 Mass Spectrometry and Protein Analysis

The identification of proteins and to determine their covalent structures has been playing a fundamental role in the life sciences since the amino acid sequence of proteins provides a link between cell physiology and genetics indirectly, which would provide a window into complex cellular regulatory networks [80].

Significant progress has been made in the past several decades toward protein analysis research. Before the genomics revolution, chemical or enzymatic methods were used to probe the covalent structure proteins via detecting the products of such reactions were by fluorescent or ultraviolet (UV) absorbance spectroscopy. For example, polypeptides were sequenced by Edman degradation by stepwise chemical degradation from the N terminus to the C terminus through subsequent identification of the released amino acid derivatives by UV absorbance spectroscopy. In the past several decades, with the increased performance and versatility of the instrumentation, mass spectrometry is the central element which creates detectors providing superior information. Although the trend toward mass spectrometry as the technique of choice for identifying the complicated structure of proteins was accelerated by the genome project that demonstrated the power of high-throughput and comprehensive analyses of biological systems, the analysis of a full proteome presents a formidable task and remains to be achieved for any species. This task is challenging because proteomes have a large and unknown complexity in which proteome exceeds by far the number of genes in the corresponding genome in a species. This diversity origins from the fact that a particular gene can generate multiple distinct proteins by the diverse protein processing mechanisms of alternative splicing of primary transcripts, the presence of sequence polymorphisms and posttranslational modifications [80, 81].

The applications of mass spectrometry on large biopolymers and biomolecules such as

proteins has drawn a lot of interest in developing MS-based techniques that would extend the research on proteomes studies. Comparing to other different physical principles of ions form the foundation for various types of mass analyzers, mass spectrometry is unique among the proteomes analytical instrumental methods. These specificities are mass-to-charge ratio determined based on path stability in a quadrupole mass filter mass analyzer, trajectories in a magnetic sector mass analyzer, velocity in a time-of-flight mass analyzer, orbital frequency in an ion cyclotron resonance mass analyzer.

Without exception, there are advantages and disadvantages for each mass analyzer [82]. The first ion-trapping device was built in the early 1920s by Kingdon using only electrostatic field based on a straight wire along the axis of the outer cylindrical electrode. Later Knight advanced the shapes and design of the electrode and made this trap design to be called ‘ideal Kingdon trap’. In 1999, Makarov invented a new type of mass analyzer-the Orbitrap. However, the Orbitrap mass analyzer was not commercialized until 2005 by Thermo Fisher Scientific [82]. It is a relatively new mass analyzer compared with other types but quickly dominates the market owing to its higher mass accuracy and mass resolution compared to linear ion trap. Last but least its price is much lower than Fourier Transform Ion Cyclotron Resonance (FT-ICR). Also using the electrostatic field, the ion-trapping in Orbitrap specially shaped electrodes compared with Kingdon trap as follows: It is equipped with a spindle-like central electrode and a barrel-like outer electrode (Figure 1.8).

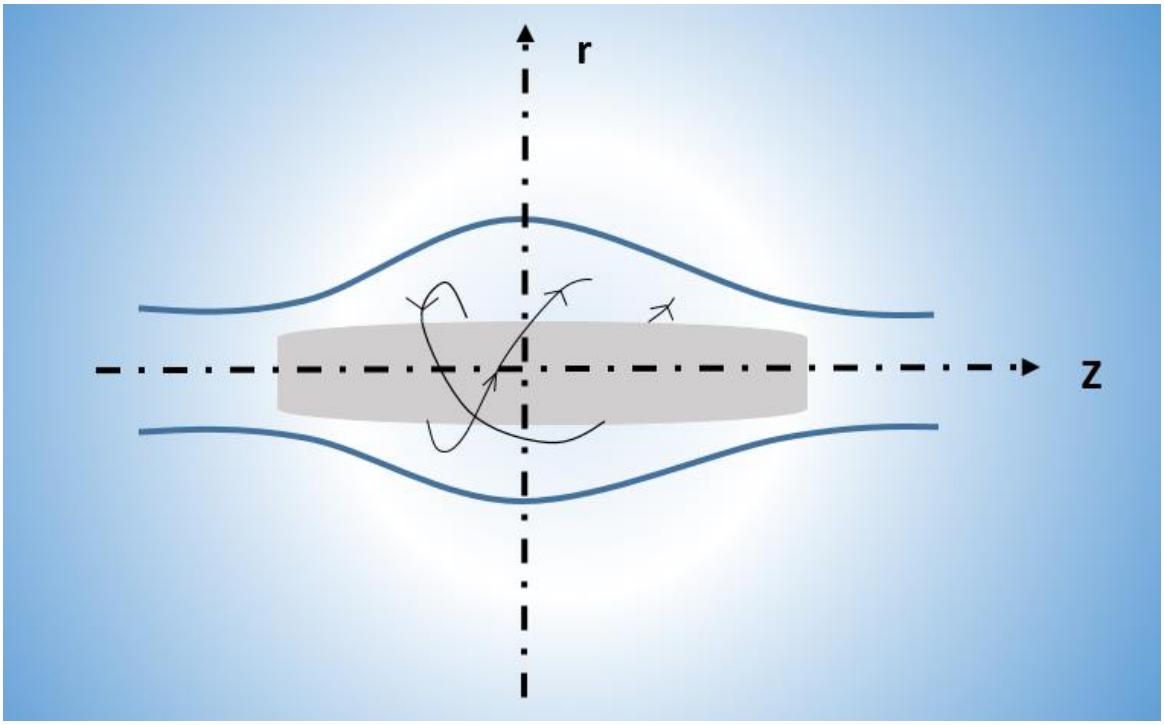

Figure 1.8 Simplified Schematics of an Orbitrap.

The ions trapped in the Orbitrap can be described with potential distribution:

$$U(r,z) = \frac{k}{2}(z^2 - \frac{r^2}{2}) + \frac{k}{2}(R_m)^2 \ln[\frac{r}{R_m}] + C \quad (1.1)$$

Where *k* is field curvature, z and r are cylindrical coordinates, C is a constant, and R_m is the characteristic radius. This field is quadro-logarithmic because it is a sum of a quadrupole field of the ion trap and a logarithmic field of the cylindrical capacitor [83]. In the field (1.1), ions move in an intricate spiral trajectory. Three characteristic frequencies of the ion can be derived by solving the equation of motion in polar coordinates (*r*, j, *z*) for ions with mass-to-charge ratio *m/q*,

$$\omega = \sqrt{\left(\frac{q}{m}\right)k} \quad (1.2)$$

where ω is the frequency of axial oscillation[84]

$$\omega_r = \omega\sqrt{\frac{(\frac{R_m}{R})^2 - 1}{2}} \quad (1.3)$$

where ω_r is the frequency of radial oscillations[84]

$$\omega_{\varphi} = \omega\sqrt{\frac{(\frac{R_m}{R})^2-1}{2}} \quad (1.4)$$

where ω_{φ} is the frequency of rotation [83].

Out of the three characteristic frequencies, only the frequency of the harmonic axial oscillation ω is inversely proportional to the square root of the mass-to charge ratio *m/q* of the ions and completely independent of the position and energy of ions. Based on this specificity, this frequency can be used for mass analysis.

The platform of image current detection used on Orbitrap relies on that an ion cloud attract with the form of positive ions or repel with the form of negative ions the electrons of the detection electrodes when it passes divided outer electrodes. The image current signal can then be transformed and translated into the frequency domain by Fourier Transform, resulting in an accurate reading of ions' mass-to-charge ratios. Orbitrap can achieve very high mass measurement accuracy which can be measured with 3 ppm.

The time of analysis increases proportionately with the resolution required as the harmonic axial oscillation frequency of each *m/z* is translated from image current. Other than analysis time, ultrahigh vacuum ($<2x10^{-10}$ mbar) is also critical for high resolving power in order to allow ions to travel sufficiently long transients in Orbitrap. However, ion collides with background gas molecules, which result in loss of phase coherence of ion packet. This can lead to a decrease in the signal intensity of the transient [85].

The Orbitrap mass spectrometer used in this dissertation is Thermo Scientific QExactive Orbitrap. The schematic of the mass spectrometer is shown in Figure 1.9. QExactive Orbitrap consists of several components: S-lens, ESI ion source, transfer octopole, quadrupole mass filter, a curve RF-only quadrupole, Orbitrap mass analyzer and high-energy collision cell (HCD). As shown in the schematic, ions generated at the ion sources need to pass through a series of components before analyzed in the Orbitrap. C-trap is used to accumulate and thermalize ions, which is important for the high

performance of Orbitrap. Cooled and accumulated ions then can be injected from C-trap to Orbitrap with narrow energy distribution and injection angle [85, 86].

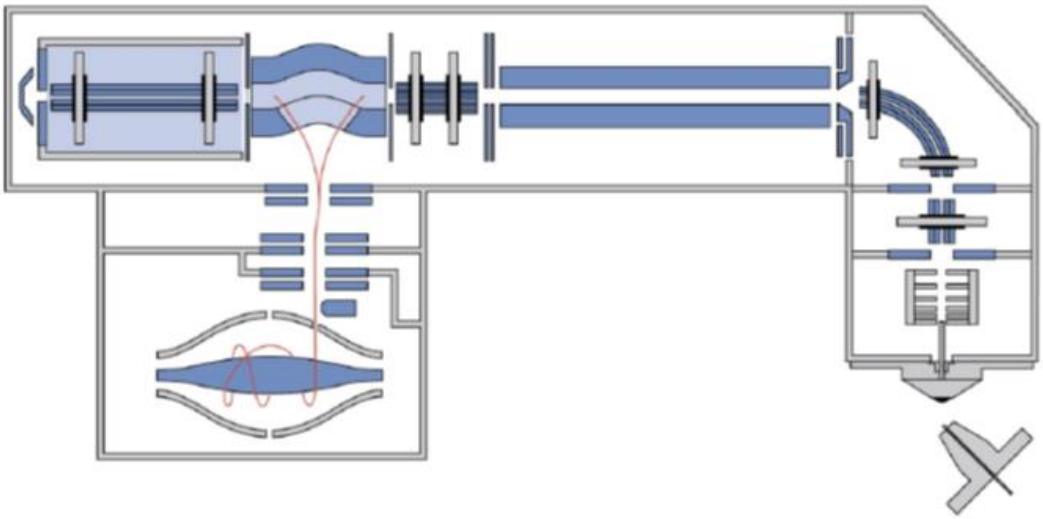

Figure 1.9 Schematic of Q-Exactive mass spectrometer (Thermo Scientific) [87].

Concerning the specific analytical requirements, The QExactive can acquire MS data at several different mass resolutions. For example, frequency is the physical property which can be measured at the highest accuracy. With the use of internal calibration ion, the mass accuracy is usually within 2 ppm. High-resolution accurate mass data with the corresponding mass error can help to achieve the defining of the elemental composition information, which is extremely useful in providing another level of information regarding identifying unknown compounds.

For quite a long time, mass spectrometry was restricted to small compounds as a consequence of the lack of effective techniques to softly ionize and transfer the ionized molecules from the condensed phase into the gas phase without excessive fragmentation [81]. Until to the late 1980s, two techniques-electrospray ionization (ESI) [88] and matrix assisted laser desorption/ionization (MALDI) [89] were developed for the routine and general formation of molecular ions of intact biomolecules, which dramatically made polypeptides accessible to mass spectrometric analysis [90].

ESI is one of most widely used ionization methods in MS owing to the following

reasons: a) wide range of eluent flow rate, b) the comparative simplicity of ESI in introducing ions from solution phase to gas phase, and c) applicability to a large number of analytes with diverse chemical and physical properties. ESI ionization takes place at atmospheric pressure, which is a 'soft-ionization' method since there is no fragmentation during the electrospray ionization process. In spite of the increasing popularity of the soft-ionization mass spectrometry methods for a variety of large synthetic and biological molecules, the analysis of proteins or peptides using different soft ionization techniques is under-developed by comparison [91].

Another 'soft-ionization' technique- MALDI coupled with time-of-flight (TOF) mass analyzer is widely used in the study of large biopolymers such as protein and polypeptides for molecular weight (MW) determination. However, the reproducibility and quality of MALDI-MS data are largely associated with sample preparation, which has led to underprivileged inter-laboratory consistency [92]. MALDI remains a valuable alternative ionization technique for peptides. Comparing to ESI, MALDI MS is more sensitive and tolerant to the presence of contaminants such as salts or small amount of detergent which may be introduced in the process of sample preparation. The MALDI technique has primarily been used in conjunction with ToF (Time-of-flight) analyzers for molecular mass determination.

In this dissertation, ESI-MS is the major analytical technique used to study and characterize the chemical compounds and protein of interest.

1.8 Discovery of Capuramycins

Capuramycin, initially called 446-S3 substance about the name of its producer strain *Streptomyces* griseus 446-S3, was discovered in a program of screening for new antibiotics in 1986 [93]. The structure of capuramycin contains three distinct moieties: a rare unsaturated hexuronic acid, uridine-5'-carboxamide (CarU), and an aminocaprolactam (LACL) (Figure 1.10). Capuramycin was proved to have excellent inhibitory activity against bacterial tranlocase I (MraY) of *Mycobacterium smegmatis*

ATCC 607 (MIC 3.13μg/mL) and *Streptococcus pneumoniae* (MIC 12.5μg/mL) [93].

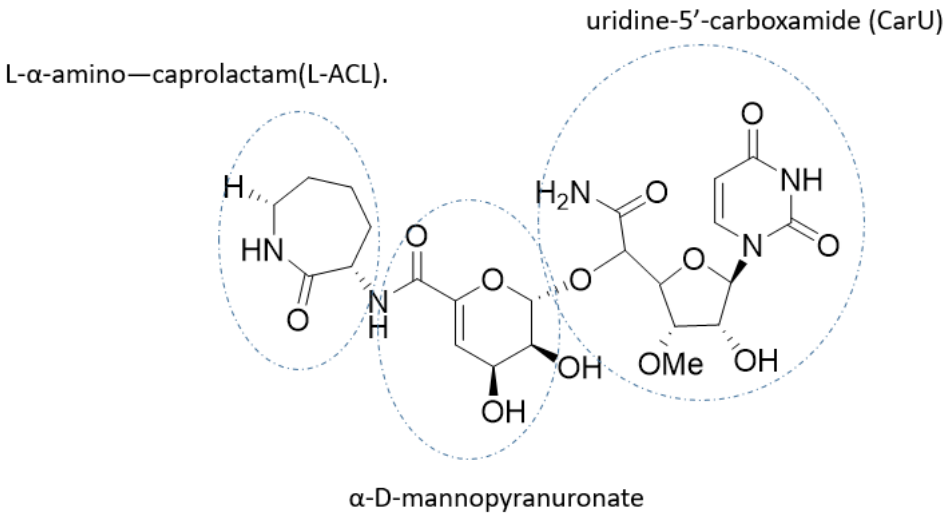

Figure 1.10 Structures of capuramycin.

Capuramycin-type antibiotics consist of A-500359s from *Streptomyces griseus* SANK 60196 [94], A-503083s from *Streptomyces sp.* SANK 62799 [56], and A-102395 from *Amycolatopsis* sp. SANK 60206 [58]. They were discovered using an activity-based screen to identify inhibitors of bacterial MraY. Yasunori elucidated 500359 A, C, D, G as 6'''-methylcapuramycin, 3'-demethyl-6'''-methylcapuramycin, 2''-deoxy-6'''-methylcapuramycin and 3'-demethylcapuramycin, respectively. The in vitro activity assay demonstrated that capuramycins showed the potent inhibitory effect on MraY. The mode of action of A-500359s was demonstrated to be the noncompetitive inhibition to undecaprenyl-phosphate [68].

Interestingly, Muramatsu et al. showed that A-500359 E showed a slightly reduced inhibitory activity (IC_{50} 27ng/mL), lacking the L-ACL moiety [95]. One year after that, another new series of antibiotics that are structurally related to capuramycin- A-503083s were discovered from *Streptomyces* sp. SANK 62799 [56]. The structures of A-503083s differ from A-500359s in the CarU moiety, which is modified with a 2'-O-carbamoyl group. Not surprisingly, A-503083s showed excellent inhibition of MraY

(e.g. IC_{50} 38 nM for A-503083 B), except A-503083F (2c) that the L-ACL structure is substituted by a carboxylic acid (IC_{50} 17.9 μM) [56].

A-102395, the latest member of the capuramycin family, was isolated from culture broth of *Amycolatopsis sp.* SANK 60206 by Ryo in 2007 [58]. It has a unique arylamine-containing polyamide instead of an ACL (Figure 1.11). Although A-102395 showed the potent inhibitory activity for bacterial translocase I with IC_{50} values of 11 nM which is as potent as that of capuramycin (IC_{50}: 18 nM), it lacks antimicrobial activity against various strains probably due to poor permeability of bacterial membrane [58]. The potency of the inhibition of translocase I by A-102395 mainly comes from the core skeleton and the other structural features.

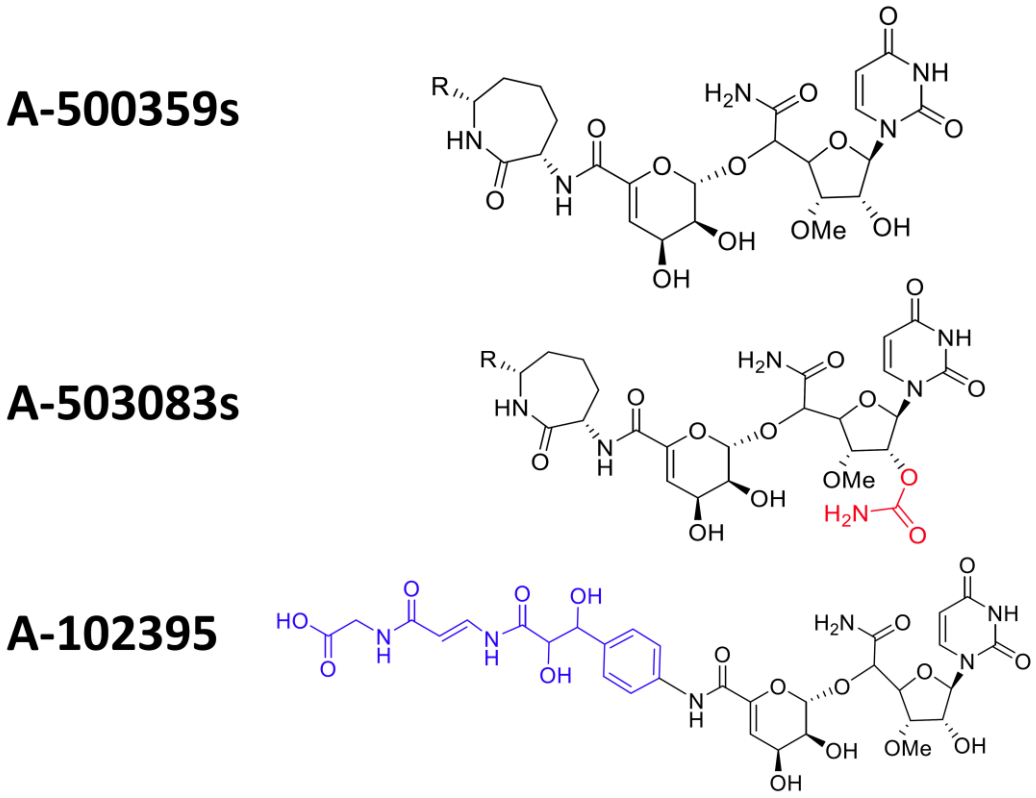

Figure 1.11 Structures of capuramycin-type antibiotics.

Due to the selective inhibitory activity against *Mycobacterium tuberculosis* and low toxicity, capuramycins are considered important leads for the development of anti-TB antibiotics. A series of semisynthetic analogues such as SQ922 and SQ641, have been

developed with better anti-TB activity by some biomedical company [96-99]. Notably, both capuramycins and their analogues were bactericidal and killed *M. tuberculosis* much faster than other first-line anti-TB drugs. This would lead to a very bright prospective in reducing the time frame for effective anti-TB chemotherapy compounds [100].

1.9 Current Understanding of the Biosynthetic Pathways of A-102395

Generally, capuramycin consists three structurally distinct components: a uridine-5′-carboxamide (CarU), an unsaturated α-D-mannopyranuronate, and an L-α-amino-caprolactam (L-ACL). The A-503083s have same structures as to those of A-500359s with one subtle exception: CarU is modified with a 2′-O-carbamoyl group, which has been turned out has minimal effect on MraY inhibition and antibacterial activity. A-102395 significantly diverges from the other capuramycins in both structures, wherein the L-ACL is substituted with an unusual polyamide arm.

Since the discovery of capuramycin in 1986, preliminary evidence of building blocks for each component of capuramycin was interrogated by using isotopic feeding experiments. Studies suggested that the unsaturated hexuronic acid is derived from D-mannose, L-ACL component originates from L-Lys, and CarU originates from D-ribose and L-Thr [68, 101-103].

Previously, the biosynthetic gene clusters for A-500359s and A-503083s have been reported. The pathway to CarU biosynthesis begins with the unusual β-hydroxy-L-amino acid GlyU as an intermediate following successive catalysis by a non-heme Fe(II), αKG: UMP dioxygenase, and an L-Thr: UA transaldolase. Cpr19/LipL leads to the generation of UA from mononucleotide. While the formation of the unsaturated α-D-mannopyranuronate is not understandable just based on bioinformatics analysis and prediction. Funabashi et al.[104] identified the gene cluster for A-500359 in *Streptomyces griseus* SANK60196. In this study, an abundantly transcribed NDP-glucose dehydratase (NGDH) gene in a high producing strain was identified and then

used as a probe to give two contiguous cosmids containing a minimum of 21 open reading frames (orfs, *orf7-orf28*). The genetic locus for A-503083 from *Streptomyces* sp was also identified by using the digoxigenin-labeled probes. Two genes in SANK 62799 *capU* and *capB* encoding a nonribosomal peptide synthetase and a carbamoylgtransferase, respectively, were also identified. This lead to the discovery of the entire gene cluster that includes 21 orfs (*capA-capW*). Additional evidence that the correct gene clusters were identified was provided by the functional characterization of several genes involved in A-503083 biosynthesis including an L-ACL transacylase (CapW), two methyltransferases CapK and CapS and a 2′-carbamoyl transferase (CapB) [101].

Bioinformatic analysis of capuramycins revealed that *orf21* has a close sequence similarity to aminoglycoside phosphotransferase(APH). The heterologous expression of *orf21* in two strains of *E. coli ΔtolC* and *Streptomyces albus* provided resistance to capuramycin [104]. Later in 2010, Yang et al. revealed that CapP [105], a gene product with high sequence similarity to ORF21, functions as an ATP-dependent phosphotransferase which transfers the γ-phosphate of ATP to the 3″ -hydroxyl of the unsaturated hexuronic acid moiety of A-503083. The 3″-phospho-capuramycins lost much antibacterial activity with the IC_{50} in the bacterial MraY inhibition assay increasing more than 200-fold compared to the wild type [105].

However, the study of biosynthesis pathway of polyamide moiety of A-102395 is quite challenging but valuable because it contains highly unusual chemical features. The 3-(4-aminophenyl)-2,3-dihydroxypropanoic acid component of the polyamide is predicted to be formed from chorismic acid and malonyl-CoA by eight ORFs *cpr12* and *cpr32–cpr38*. The proposed pathway initially proceeds by the conversion of chorismic acid to p-aminobenzoic acid, which catalyzed by Cpr38 with bidomain function involving amidohydrolysis of L-Gln with ammonia channeled and incorporated into chorismic acid to generate ADC. Another protein Cpr12 is proposed to function as ADC lyase (ADCL) to catalyze the elimination of pyruvate to form para-aminobenzoic acid

(PABA). The remaining steps require C–C bond formation between a C-2 extender unit and the carboxylic acid of activated PABA. These steps were proposed to be initiated by Cpr37-catalyzing as the adenylation domain protein and Cpr34 and Cpr35 which are predicted to work in concert to catalyze decarboxylative condensation between thioester and a malonyl-S-acyl carrier protein (ACP) during aromatic polyketide biosynthesis catalyzed by type II polyketide synthases. Cpr33 was proposed to function as a putative 3-oxoacyl-ACP reductase catalyzes reduction to the β-hydroxythioester intermediate. Cpr32 was predicted as a luciferase-like monoxygenase catalyzing hydroxylation to give 3-(4-aminophenyl)-2,3-dihydroxypropanoic acid. The remaining steps in polyamide biosynthesis involve the formation of two amide bonds was proposed to be catalyzed by a series of putative proteins encoded within ORFs *cpr47–cpr57* that are associated with amide bond-forming events and the final step of coupling of the arylamine-containing polyamide to the capuramycin core via amide bond formation [106].

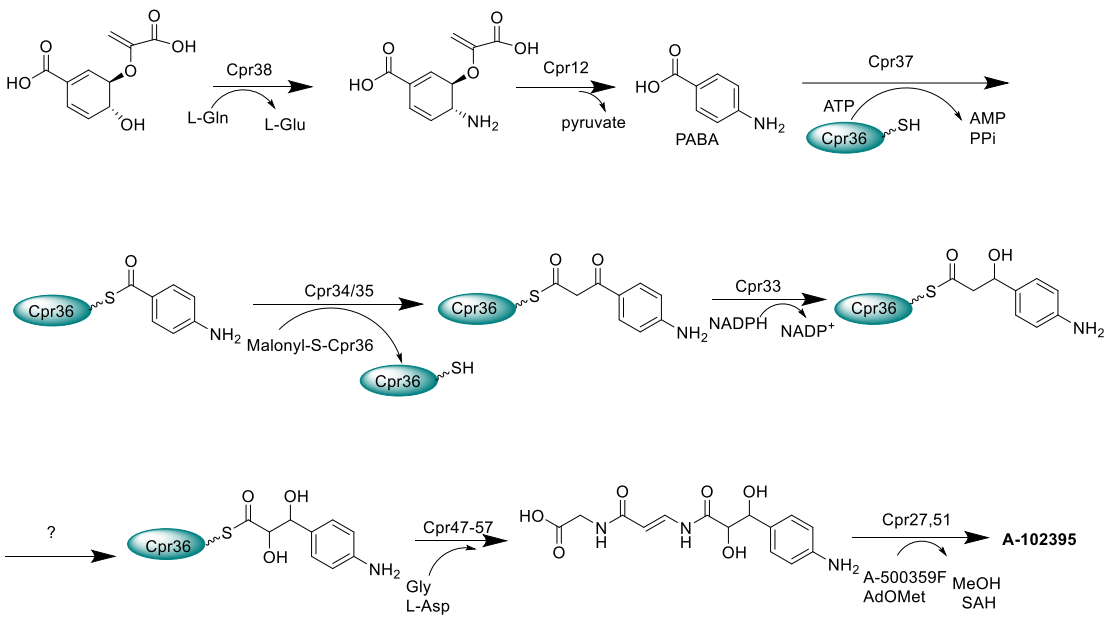

Figure 1.12 Proposed biosynthetic pathway of the unique arylamine-containing polyamide of A-102395.

1.10 Aims of this Study

The specific aim of this study was the identification and characterization of the biosynthetic gene cluster for polyamide assembly of A-102395. We set following goals:

1. Cloning and heterologous expression of unknown key enzyme for the biosynthesis pathway of A-102395 such as Cpr38 and Cpr12 that responsible 4-amino-4-deoxychorismate synthase and aminotransferase.

2. Cloning and heterologous expression of Cpr36 and Cpr37 that responsible aryl carrier protein and adenylation domain protein.

3. Investigate how C-C bond between the carboxylic acid of PABA and a C-2 extender unit was formed and the specific decarboxylative condensation mechanism.

4. Thorough investigation formation of β-hydroxythioester in capuramycin A-102395 biosynthesis metabolic pathway and the function of Cpr33 which is proposed to function as 3-oxoacyl-ACP reductase to catalyze the formation of β-hydroxythioester intermediate of A-102395.

Chapter 2: Elucidating the mechanism of Cpr38 and Cpr12

2.1 Background

Chorismic acid serves as the branch point precursor for metabolites which play a vital role in the biosynthesis of many critical aromatic products in *E. coli*. These metabolites and their end products include but are not limited to prephenate (tyrosine, phenylalanine), anthranilate (tryptophan), PABA (folic acid), p-hydroxybenzoate (ubiquinone) and isochorismate (enterochelin, menaquinone) (Figure 2.1) [107]. Chorismic acid converts to PABA which is a precursor of some crucial metabolites in both prokaryotes and eukaryotes [108]. PABA is generally synthesized in two steps from chorismate and glutamine. The first step is the conversion into 4-aminodeoxychorismate (ADC) and glutamate, respectively. The second step is the conversion between ADC and PABA.

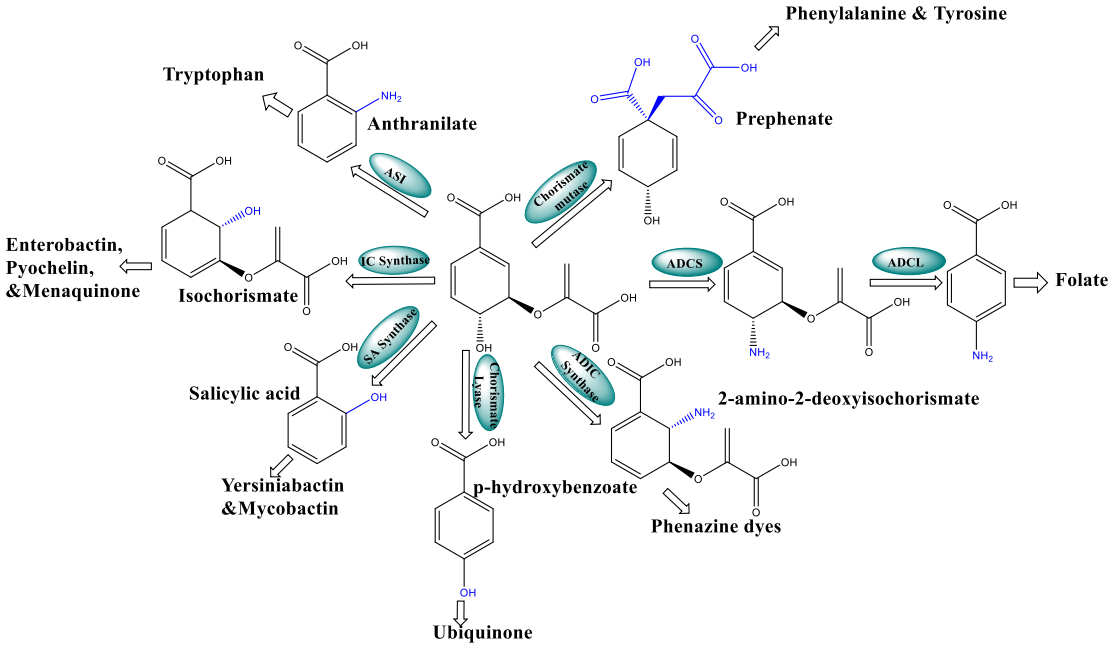

Figure 2.1 Metabolic branching points from chorismate.

The pathway of PABA moiety derived from chorismate was poorly characterized i enzymatic steps until the 1990s in the bacterial biosynthesis of folate coenzymes. This is partially because the enzyme activity in wild-type *E. coli* had been problematic to

detect in crude extracts [109-111]. PABA was proposed to be synthesized from chorismate which is needed to synthesize crucial aromatic products and their derivatives [112-114]. In 1989, Nichols et al. showed that the conversion of chorismate to PABA occurs in two separate protein-catalyzed steps, rather than one as previously thought [115]. The first step in the synthesis of PABA is the amination of chorismate to form ADC. In the second step, ADC is aromatized with loss of pyruvate to form PABA. Specifically, in *E. coli*, two proteins named PabA and PabB forming a heterodimeric complex catalyzes this reaction whereas just a single protein is involved in plants and lower eukaryotes [112, 116, 117]. The heterodimeric complex catalyzes a reaction with an ordered Bi Bi mechanism in which chorismate binds first where PabA only has similar activity with glutaminase [118, 119]. PabB alone then converts chorismate into ADC in the presence of nascent NH_3 [120]. Hossain et al. found that the *pabA* gene encodes a 21-kDa protein (PabA) with high sequence homology to the TrpG component, which was found in the biosynthesis pathway of o-aminobenzoate (anthranilate) [117]. Both *TrpG* and *pabA* encode a glutaminase activity, providing nascent NH_3 for the chorismate amination [121]. In the reaction catalyzed by PabA and PabB, glutamine is the source of NH_3 [118, 122]. The 51 kDa *pabB* gene product, has ~26% sequence identity substantial homology to the *trpE* gene product which is the large subunit of the anthranilate synthase TrpE/G complex [112, 113]. Policastro et al. found that the TrpE protein catalyzes NH_3-dependent chorismate amination to 2-amino-2-deoxyisochorismate and its subsequent aromatization by elimination of the components of pyruvate [114]. It was proved that that PabB also catalyzes similar regiospecific amination and then aromatization of ADC [123].

The crystal structure of PabB [124] has a complex α/β fold and indicates strong resemblances with the TrpE subunit of anthranilate synthase which is another tryptophan allosterically regulated chorismate-utilizing enzyme (Figure 2.2A &B) [125]. Although PabB activity is not regulated by tryptophan, a tryptophan molecule was found deeply embedded in the structure, consolidating the evidence that the idea

that PabB and TrpE are evolutionarily related [124]. In recent times, PabB was identified as an interesting target for discovery and development of novel antimicrobial reagents with an example of (6S)-6-fluoroshikimate owing to its ability of conversion into 2-fluorochorismate, which had been proved to strongly inhibit PabB [126]. Genomic data shows that lower eukaryotes also have a bi-domains similar to above mentioned PabA and PabB and separately positioned at the C- and N-terminal part of the enzyme [127-129].

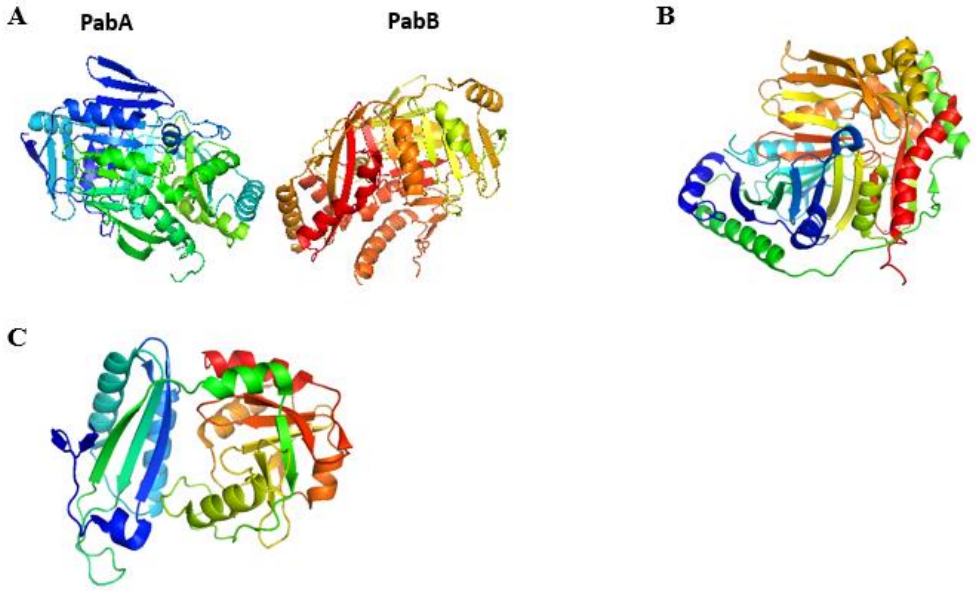

Figure 2.2 The crystal structure of chorismate pathway related enzymes. 4-Amino-4-Deoxychorismate Synthase(A), TrpE subunit of 4- Anthranilate Synthase(B) and Amino-4-Deoxychorismate Lyase(C) from *E. coli*. (figures were redrawn with pymol using published data, PDB ID: 1K0E, PDB ID:5CWA PDB ID:1I2K).

In 1990, Ye et al. suggested that there is the third enzyme in PABA biosynthesis called aminodeoxychorismate lyase (ADCL) encoded by the gene of *pabC* based on the data of structure of an intermediate in-ammonia-dependent PABA biosynthesis from chorismate. ADCL was proposed to catalyze the reaction of elimination of pyruvate to form PABA. This assumption was entirely contrast to anthranilate synthase where no additional proteins are involved in the biosynthesis pathway of chorismate to

anthranilate. Additionally, the biosynthesis of PABA differs from that of anthranilate. In ADC, there are two subunits of synthase encoded by p-aminobenzoate biosynthetic genes *pabA* and *pabB* [130] whereas all the traditional *E. coli* tryptophan biosynthetic genes are located in just a single operon. In 1992, Jacalyn et al. found that ADCL encoded by *E. coli pabC* contains a pyridoxal phosphate cofactor. A strain that was deficient in *pabC* exhibited a PABA requirement [131]. The ADCL activity in plants was supported by a second enzyme targeted to plastids which was the unique site of PABA synthesis in plants [132].

The crystal structure of ADCL shows it contains a tightly bound pyridoxal 5'- phosphate (PLP) as a cofactor. A lot of PLP-dependent enzymes are classified as type I based on similarities in sequence, but ADCL is placed in fold type IV which transfer protons on the re-face of cofactor which was confirmed by X-ray [131]. While in most PLP-dependent enzymes, the catalytic residue lysine shuttles protons on the si-face of the planar π-system of the substrate-cofactor complex [133]. In turn, ADCL is expected to have a conventional structure as other type IV PLP-dependent enzymes such as DAAT and BCAT because the configuration between the cofactor and catalytic base at the active site is considered to reflect the stereospecificity of the hydrogen transfer [133].

In this chapter, we will discuss the identification and characterization of two crucial biosynthetic gene cluster for A-102395: Cpr38 and Cpr12, which was proposed to function as the 4-amino-4-deoxychorismate synthase (ADCS) and 4-amino-4-deoxychorismate lyase ADCL in the biosynthesis pathway of A-102395 by bioinformatic analysis.

2.2 Materials and Methods

2.2.1 Instrumentation, Chemicals and Reagents

PLP, bicine, chorismic acid from *Enterobacter aerogenes*, L-glutamate, L-aspartate and organic solvents, and media were purchased from Sigma-Aldrich (St. Louis, MO) or Promega (Madison, WI). All other reagents used were analytical grade. Buffers, salts,

and media components were purchased from Fisher Scientific (Pittsburgh, PA). Synthetic oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA). Wizard® Plus SV Minipreps DNA Purification Systems, Wizard® SV Gel and PCR Clean-Up System were purchased from Promega (Madison, WI, USA). Restriction enzymes, DNA ligase, and Vent DNA polymerase were from New England Biolab. Taq DNA polymerase was from Promega. PCR was performed with a Veriti® 96-well thermal cycler from Applied Biosystems. Nanodrop 2000 UV-Vis spectrophotometer (Thermo Scientific, Waltham, MA) was used to measure concentration and purity of DNA, RNA and proteins.Synthetic oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA). UV/Vis spectroscopy was performed with a Bio-Tek μQuant microplate reader using Microtest™ 96-well plates (BD Biosciences) or a Shimadzu UV/Vis-1800 Spectrophotometer equipped with a TCC-240A thermoelectrically temperature controlled cell holder. HPLC was performed with an Agilent 1200 separation module (Agilent Technologies, Santa Clara, CA) with an analytical Agilent C-18 column (150 mm x 4.6 mm, 5 μm). LC-electrospray ionization (ESI)-mass spectroscopy (MS) was performed using an Agilent 6120 Quadrupole MSD mass spectrometer equipped with an Agilent 1200 Series Quaternary LC system and an Eclipse XDB-C18 column (150mm x 4.6 mm, 5 μm, 80Å). Ni-NTA agarose was purchased from Qiagen (Valencia, CA). Amicon Ultra 10000 MWCO centrifugal filter was purchased from Millipore (Billerica, MA). PD-10 desalting column was purchased from GE Healthcare (Pittsburgh, PA). DNA sequencing was performed using the BigDye™Terminator version 3.1 Cycle Sequencing kit from Applied Biosystems, Inc. (Foster City, CA) and analyzed at the University of Kentucky Advanced Genetic Technologies Center. Nova Blue Giga Singles Competent cells, BL21 (DE3) competent cells were purchased from Invitrogen (Camarillo, CA). Takara LA taq DNA polymerase was purchased from Takara Bio Inc (Otsu, Shiga, Japan). Restriction enzymes were purchased from New England Biolabs (Ipswich, MA). Phusion Green Hot Start II High-Fidelity DNA Polymerase was

purchased from Thermo Scientific (Waltham, MA). A-102395 strain and standard was a gift from Dr. Koichi Nonaka, Daiichi-Sankyo.

2.2.2 Introduction of pNCap01 and pNCap03 into *E. coli* BW25113/pIJ790 Strain

The supercosmids pNCap01 and pNCap03, a gift from Dr. Koichi Nonaka at Daiichi Sankyo. To ensure the accuracy, the supercosmids were introduced and amplified by electroporation as described below. Supercosmids DNA was repurified and collected to microcentrifuge tubes as the protocol of Wizard® SV Gel and PCR Clean-Up System DNA kit. To determine transformation efficiency, 1 μL of the pUC19 control DNA was added to a microcentrifuge tube. The DNA was precipitated with ethanol and resuspend in TE buffer (10 mM Tris HCl, pH 7.5; 1 mM EDTA). ElectroMAX™ DH5α-E™ cells were thawed on wet ice and the cells were mixed by tapping gently and then added into a pre-chilled microcentrifuge tube. Unused cells were refrozen the in a dry ice/ethanol bath for 5 minutes before returning them to the -80°C freezer. The cell/DNA mixture was added into a chilled 0.1 cm cuvette and electroporate with the following condition 2.0 kV, 200 Ω, 25 μF. 1 mL of SOC medium was added into the cells in the cuvette and then transferred to a 15 mL snap-cap tube. The tube was shaken at 225 rpm (37°C) for 1 hour. Cells were spread on prewarmed LB plates containing 100 μg/mL ampicillin and incubate plates overnight at 37°C. Until colony grown up, a single and well-isolated colony from a fresh LB agar plate (containing 100 μg/mL ampicillin) was inoculated into 10 mL of LB medium (containing the same antibiotic) and incubated overnight (18 hours) at 37°C in a shaking incubator to optimize the supercosmids.

10 mL of bacterial culture was harvested by centrifugation for 5 minutes at 10,000 x g by a tabletop centrifuge. After sucking out the supernatant, 250 μL of Cell Resuspension Solution was added into the tube and the cell pellet was completely resuspend by vortexing or pipetting. 250 μL of cell lysis Solution was added and mixed by inverting the tube four times. 10 μL of Alkaline Protease Solution was added into the tube until the cell suspension clears. The tube was mixed by inverting the tube four times and

incubated for 5 minutes at room temperature. 350 μL of neutralization solution was added. Then the bacterial lysate was centriguged at the speed of 14,000×g in a microcentrifuge for 10 minutes at room temperature. The remaining cleared lysate was transferred into the prepared spin column by decanting. The supernatant was centrifuged at the speed of 14,000×g in a microcentrifuge for 1 minute at room temperature. 750 μL of column wash solution (previously diluted with 95% ethanol) was added into the spin column and centrifuged at maximum speed in a microcentrifuge for 1 minute at room temperature. The wash procedure was repeated using 250 μL of column wash solution. The spin column was centrifuged in a microcentrifuge for 2 minutes and transferred into to a sterile new 1.5 mL microcentrifuge tube. The supercosmids DNA was eluted by adding 100 μL of nuclease-free water to the spin column and centrifuged at the speed of 14,000×g for 1 minute at room temperature in a microcentrifuge and removed the assembly from the 1.5 mL microcentrifuge tube. The supercosmids DNA was stored at –20°C for further use.

2.2.3 Cloning, Overexpression and Purification of *cpr12* and *cpr38* Genes

Gene products of Cpr12 and Cpr38 were difficult to obtain because they are insoluble using multiple hosts and a variety of expression conditions. To get these gene products, Cpr12 and Cpr38 were produced in *E. coli* BL21 (DE3) as a maltose-binding protein (MBP)-fusion to obtain soluble protein for *in vitro* assays. PCR was conducted using primers for amplification of cpr12 with *E. coli* BW- pNCap01 and cpr38 with *E. coli* BW-pNCap03 as the templates (Table 2.1).

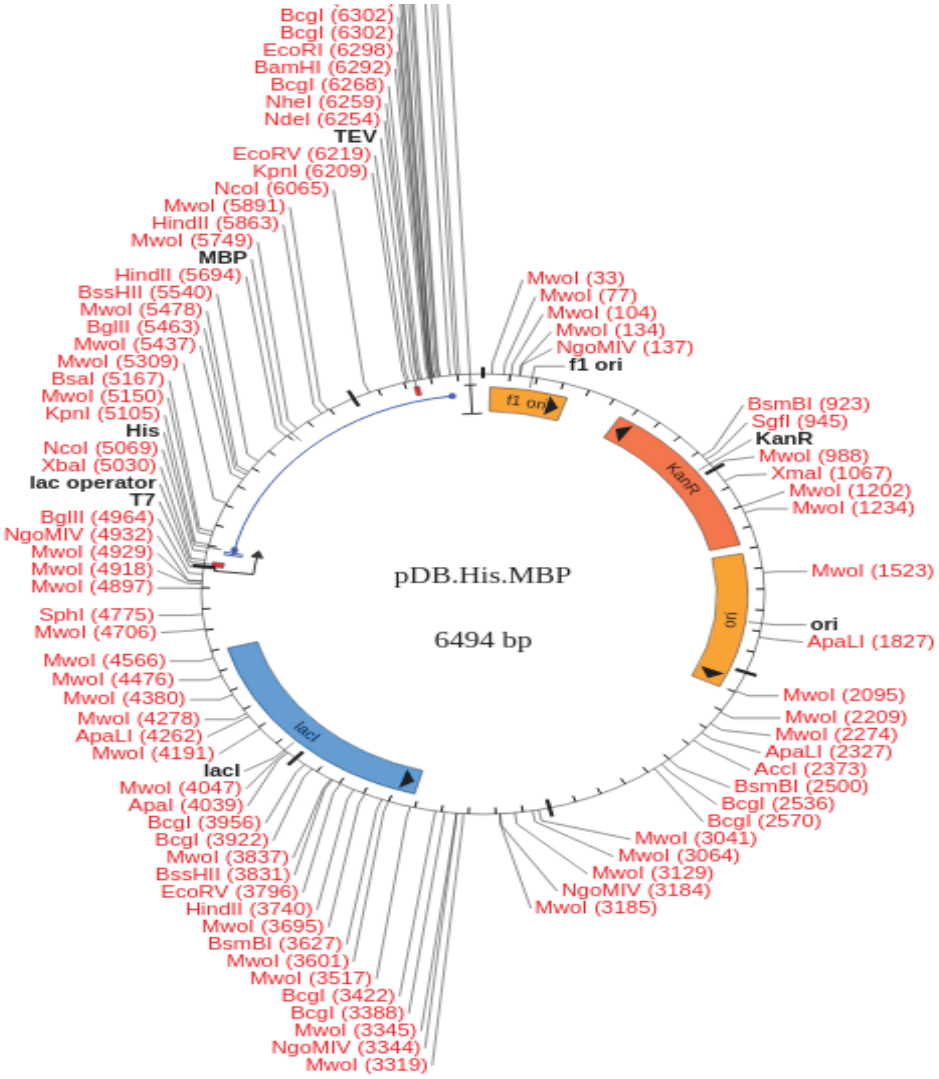

Figure 2.3 Map of pDB.His.MBP plasmid. The plasmid is resistant to kanamycin [134].

Table 2.1 Primers for amplification of *cpr12* and *cpr38*.

Name	Sequence
Cpr12_forward	5′-GAT ACGCAT ATG ATGAGGGATCCATACTGGCT-3′
Cpr12_reverse	5′-TAGAGTGAATTCTCATTGACCTGTTTCATGATC-3′
Cpr38_forward	5′-GATACGGGATCCATGACGGGTGTCGCTCCTGT-3′
Cpr38_reverse	5′-TAGAGTGAATTCTTAACAATCATTTCTCTCCA-3′

The PCR programs began at 97 °C for 2 min, followed by 20 cycles of 97 °C for 45 s, 50°C for 45 s and 72°C for 90 s, and final extension at 72 °C for 7 min for *cpr12* and 97 °C for 2 min. And the program was followed by 20 cycles of 97 °C for 45 s, 55°C for 45 s and 72°C for 90 s, and final extension at 72 °C for 7 min. The resulting linear PCR products were analyzed by agarose gel electrophoresis to confirm the correct gene size and purified by using PCR Clean-Up System DNA kit purchased from Promega (Madison, WI, USA) following the protocol. The obtained DNA and Vector pDB.His.MBP vector were doule digested as following conditions: 1 µg DNA, 1 µL of NdeI and EcorI restriction Enzyme for *cpr12* and 1 µL of BamHI for and EcorI restriction Enzyme for *cpr38*, 3 µL 10x Buffer, 3 µL 10x BSA, x µL ddH_2O varies each time to bring total volume to 30µL. The reaction systems were mixed gently by pipetting and incubated tube at 37 °C for 1 hour by adding all reagents by always following the manufacturer's instructions. To visualize the results, gel electrophoresis was conducted. The digested DNA were further isolated and purified by using PCR Clean-Up System DNA kit following the protocol above mentioned. DNA ligation was conducted to fuse the inserts to pDB.His.MBP vector (Figure 2.3) recipient plasmid by combining the following in a PCR or Eppendorf tube: 25ng Vector DNA, 75ng Insert

DNA, Ligase Buffer (1 μL/10 μL reaction for 10X buffer, and 2 μL/10 μL reaction for 5X buffer), 0.5 μL T4 DNA Ligase, ddH_2O to a total of 10 μL. The ligation reaction system was incubated at 16°C overnight by following the manufacturer's instructions. Then transformation was proceeded according to the following process: competent cells was initially taken out from -80°C and thawed on ice approximately 30 mins. Agar plates containing kanamycin antibiotic were taken out from storage at 4°C and warmed up to room temperature and then incubated in 37°C incubator. 5 μL of DNA (100 ng) was added into 50 μL of competent cells in a falcon tube. The tube was gently mixed by flicking the bottom with finger a few times. The mixture of competent cell/DNA was incubated on ice for 30 mins. Each transformation tube was placed into a 42°C water bath for 90 secs. The tubes were put back on ice for 2 min. 600 μL SOC media were added (without antibiotic) to the bacteria and grow in 37°C shaking incubator for 45 min. The bacteria cells were spun down at the speed of 14,000×g for 5 minutes at room temperature. The supernatant was discarded and remained pellet was plated onto a 10 cm LB agar plate containing the kanamycin antibiotic. Finally, individual bacterial colonies were picked and checked for successful ligations. Depending on the number of background colonies on control plate (the more background, the more colonies will need to be picked), 3-10 colonies were picked and cultured overnight for DNA purification. The DNA purification was conducted by using the Wizard® Plus SV Minipreps DNA Purification Systems following the manufacturer's instructions. After purifying the DNA, a restriction digest of 300ng of purified DNA with each enzyme used for the cloning was conducted for diagnostic purpose. The digested DNA electrophoresis was conducted to check whether two bands exist. One band should match the size of the vector and one should match the size of the new insert. Finally, constructed plasmids were verified by sequencing.

2.2.4 Overexpression and Purification of Cpr12 and Cpr38

Constructed pETMBP_1a inserted with cpr38 were transformed into *E. coli* BL21 (DE3)

cells by following the transformation protocol previously mentioned. Cells harboring pETMBP_1a-cpr38 construct in the single colony were picked up and incubated in 5 mL LB media with the 50µg/mL kanamycin at 37 °C overnight. On the second day, the culture was transferred to big flask filled with 500 mL medium of TB until the OD600 reached 0.5 when expression was induced with 0.5 mM IPTG, and the incubation was continued for another 12h at 18°C. Cells were harvested by centrifugation and resuspended in lysis buffer (100 mM KH_2PO_4, 300 mM NaCl, 1 mM EDTA, 10 mM Imidazole). The cells were chilled on ice for 15 min and then disrupted by sonication. Cell debris was pelleted by centrifugation at 15,000 rpm for 60 min at 4 °C. The lysate was filtered to exclude pellet contamination and loaded onto an affinity chromatography with Ni-NTA agarose and washed to exclude the impurities proteins by washing buffer (100 mM KH_2PO_4, 300 mM NaCl, 1 mM EDTA, 20 mM Imidazole). The recombinant proteins were eluted with 20 mL of elution buffer(100 mM KH_2PO_4, 300 mM NaCl, 1 mM EDTA, 200 mM Imidazole) and then concentrated with an Amicon Ultra 10,000 MWCO centrifugal filter (Millipore), followed by desalting of the protein using the desalting buffer containing 50mM KH_2PO_4 and 150mM NaCl. The proteins were stored in 30% glycerol at -20 °C until use. Concentration was determined by Nano Drop according to the predicted extinction coefficient of each protein, and the size and purity of recombinant proteins were assessed by 12% acrylamide SDS-PAGE.

BL21 (DE3) Gold cells harboring the pETMBP_1a-cpr12 construct were inoculated with a single colony and were grown in 5 mL LB at 37 °C overnight and transferred to 500 mL of TB media with the 50µg/mL kanamycin. IPTG was added to final concentration of 0.5 mM when the OD600 reached 6-7. Cells were allowed to grow at 18 °C and 250 rpm for an additional 20 h after induction. Cells were harvested, and the pellets were resuspended in starting buffer (100 mM KH_2PO_4, 300 mM NaCl, 1 mM EDTA, 20 mM Imidazole, 20 µM PLP). The cells were lysed and the supernatant was loaded onto the nickel resin as above mentioned. The protein was eluted with elution

buffer (100 mM KH_2PO_4, 300 mM NaCl, 1 mM EDTA, 200 mM Imidazole, 20 µM PLP) and followed by desalting of the protein using the desalting buffer containing 50 mM KH_2PO_4, 150 mM NaCl, 20 µM PLP. The identity of the enzyme was confirmed by SDS-PAGE.

2.2.5 Functional Characterization of Cpr12 and Cpr38

A standard reaction mixture LC-MS analysis contains 2mM chorismate, 200 mM bicine, pH 8.5, either 40 mM glutamine or 200 mM NH_4Cl (or other ammonium salts), 5 mM $MgCl_2$ and 20 µM Cpr38. After 6 h at 30 °C, the pH was adjusted to around 4, the sample were ultrafiltered to remove protein. 20 µM PLP and 20 µM Cpr12 were added into the sample. After 6 h at 30 °C, the sample were ultrafiltered to remove protein and injected onto the HPLC. HPLC was performed using a C-18 reversed-phase column (Agilent) under following conditions: a series of linear gradients from 0.1% TFA in 100% water (A) to 0.1% TFA in acetonitrile (B) in the following manner (beginning time and ending time with linear increase to % B): 0-16 min, 20% B; 16-17 min, 1% B; 17-21min; The flow rate was kept constant at 0.4 mL/min, and elution was monitored at 254 nm. LC-MS analysis was performed using an Eclipse XDB-C18 reversed-phase column (Agilent) under a series of linear gradients from 0.1% formic acid in H_2O to 0.1% formic acid in acetonitrile over 20 min. The flow rate was kept constant at 0.4 mL/min, and elution was monitored at 254 nm. Peaks were identified by comparing retention times with authentic standards. Positive ions were detected. The parameter setup was: gain, 10L; fragmentor voltage, 40 V; capillary voltage, 3500 V; drying gas flow,10 L/min; nebulizer pressure, 25 psi; drying gas temperature, 350 °C.

2.3 Results and Discussion

Bioinformatic analysis of secondary structure-based prediction using HHPred suggested that Cpr38 has a sequence similarity to ADCS, which was previously

assigned to catalyze a two-step reaction involving amidohydrolysis of L-Gln and incorporation of the free ammonia into chorismic acid to generate ADC [120]. Following analysis revealed that Cpr12 has a sequence similarity to ADCL (4-amino-4-deoxychorismate lyase) which catalyze the formation of PABA from ADC through elimination of pyruvate [102]. In this study, we proposed the scheme for the initiation of assembly of polyamide chain of A-102395 (Figure 2.4). Different from ADCS and ADCL from *E. coli*, the gene product of *cpr38* and *cpr12* was not soluble in the supernatant when expressed using pET28/30 vectors (data not shown).

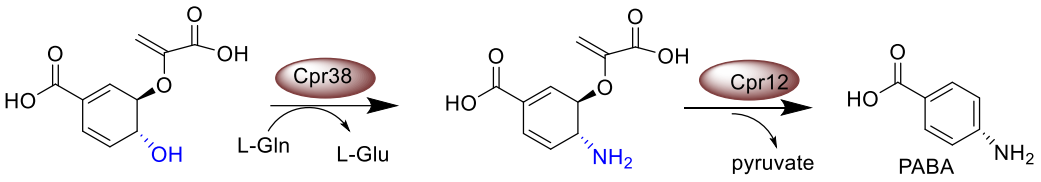

Figure 2.4 Proposed biosynthetic pathway towards initiation of assembly of polyamide side chain of A-102395.

The *cpr38* and *cpr12* genes were cloned and expressed in *E. coli* BL21 (DE3) as a maltose-binding protein (MBP)-fusion to obtain soluble protein with the expected sizes (Figure 2.5) for in vitro assays.

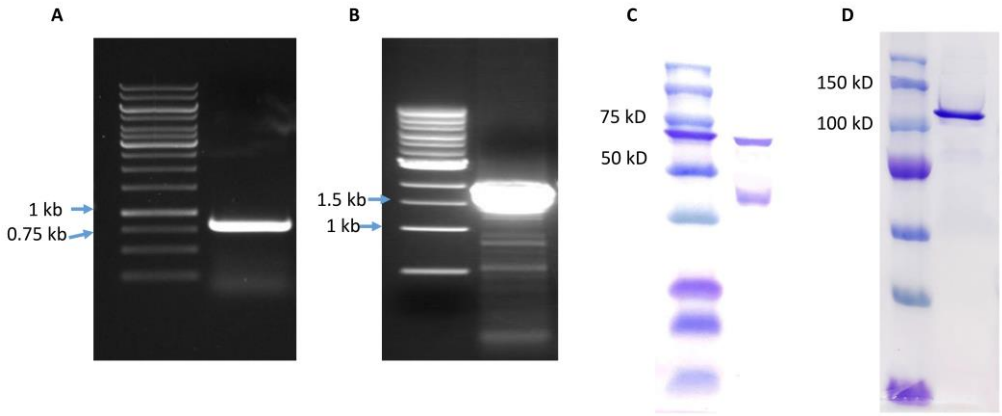

Figure 2.5 Overexpression and purification of Cpr38 of Cpr12. (A) PCR product of *cpr12*. (B) PCR product of *cpr38*. (C) SDS analysis of recombinant Cpr12 (expected MW of 71 kD). (D) SDS analysis of recombinant Cpr38 (expected MW of 115 kD).

The reactions catalyzed by Cpr38 and Cpr12 were conducted as above mentioned and subjected to chromatographic separation to identify the products formed in the reactions (Figure 2.6). As expected, LC-MS analyses of the reaction catalyzed by Cpr38 revealed $(M+H)^+$ ion at *m/z* = 226.1. The observed mass corresponded well to the predicted mass of 226.1 amu for the cationic form of ADC, in which ADCS function of Cpr38 was revealed. Once Cpr12 was added to this reaction mixture, the ADC is completely converted to a new peak which was confirmed as PABA by comparing to the PABA standard with the exact same retention time and molecular weight of $(M+H)^+$ ion at *m/z* = 138.1 (Figure 2.6 and 2.7), in which ADCL function of Cpr12 was confirmed. Negative control consists of the reaction mixture without the enzymes corresponding to the identity of chorismate (Figure 2.6B) (expected $(M+H)^+$ ion at *m/z* = 227.1, found: 227.2).

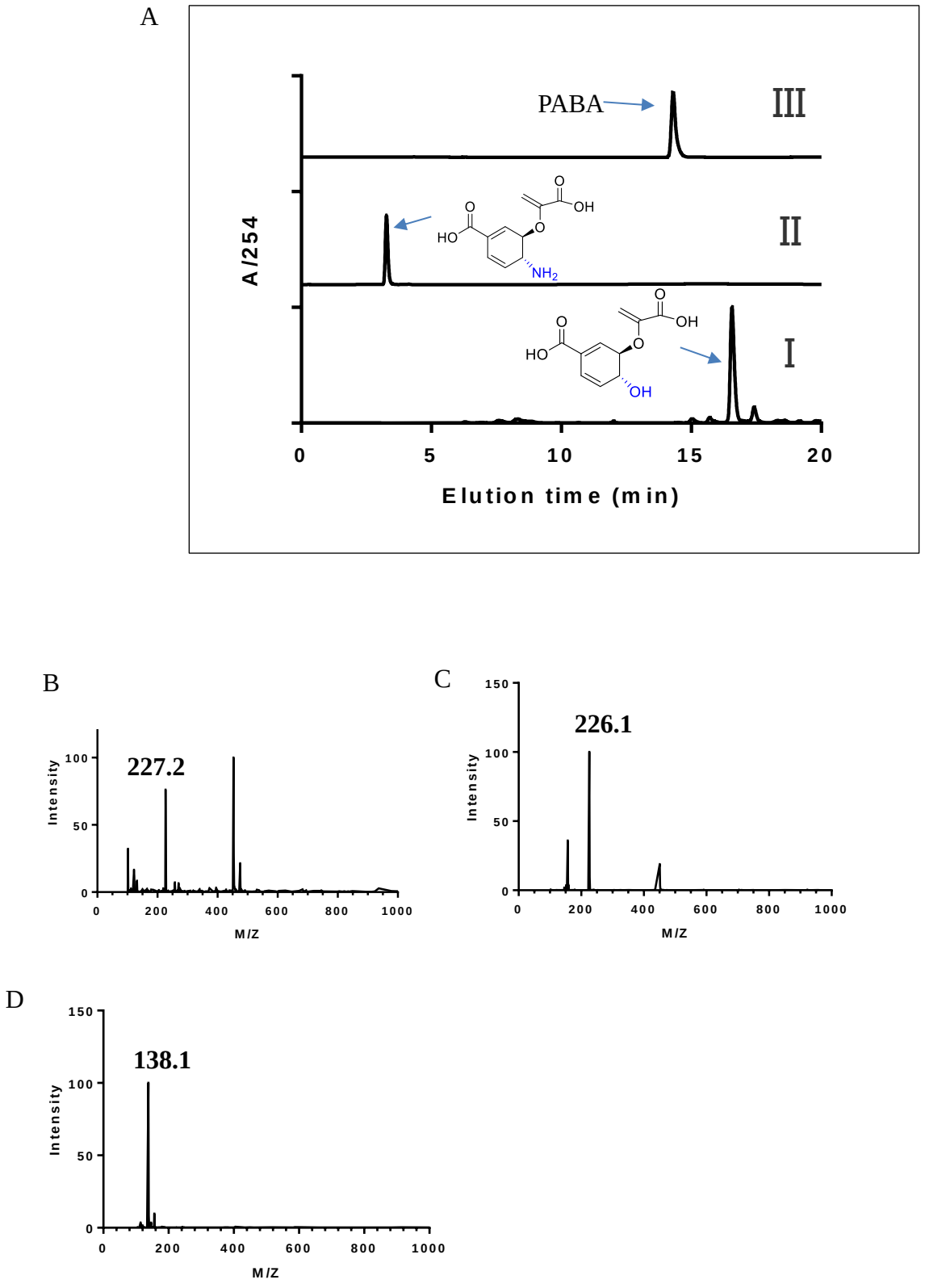

Figure 2.6 LC-MS analysis of Cpr38 and Cpr12 catalyzed reactions. (A) Extracted ion chromatograms of the reactions catalyzed by Cpr38 and Cpr12. I, 6 h reaction w/o any enzyme; II, 6 h reaction w/Cpr38; III, 6 h reaction with Cpr38 and Cpr12. (B) Mass spectrum for the ion peak eluting at time t = 16.56 min of reaction of I (chorismic acid

expected $(M+H)^+$ ion at *m/z* = 227.1, found: 227.2). (C) Mass spectrum for the ion peak eluting at time t = 3.28 min of reaction of II (ADC expected $(M+H)^+$ ion at *m/z* = 226.1, found: 226.1). (D) Mass spectrum for the ion peak eluting at time t = 14.30 min of reaction of III, (PABA expected $(M+H)^+$ ion at *m/z* = 138.1, found: 138.1).

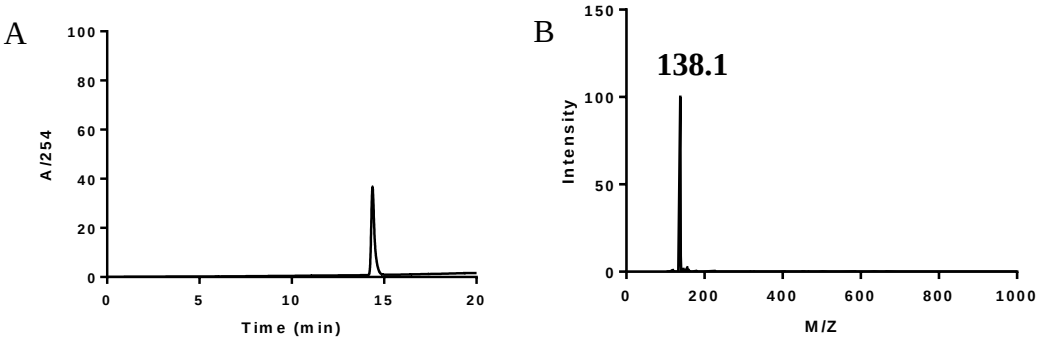

Figure 2.7 LC-MS analysis of PABA standard. (A) Extracted ion chromatograms of PABA. (B) Mass spectrum for the ion peak eluting at time t = 14.30 min of PABA (PABA expected $(M+H)^+$ ion at *m/z* = 138.1, found: 138.1).

We also introduced different ammonia resources of ammonium salts with the same concentration. These salts are NH_4Cl, NH_4HCO_3 and NH_4OAC. It revealed that the ammonium chloride had the highest product yield and NH_4OAC has the lowest, which based on the amount ratio of product and reagent after incubating with Cpr38 at 30°C for 6 hours. This is possibly because there are differences of K_d between different ammonium salts (Figure 2.8 and Table 2.2).

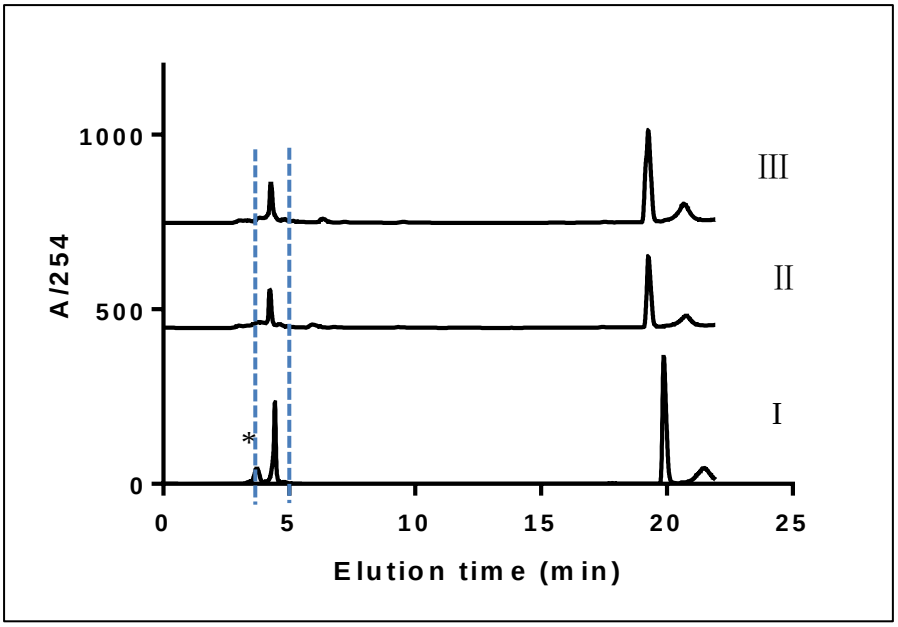

Figure 2.8 HPLC fractionation of reaction of Cpr38 with different ammonium sources. 6 h Reaction catalyzed by Cpr38 with (I) NH_4Cl, (Ⅱ)NH_4HCO_3 and (Ⅲ)NH_4OAC. The small variation in chorismate retention time is due to slightly different chromatographic conditions on different dates.

Table 2.2 Different conversion efficiency by different ammonia salt sources.

Ammonia Source (300 mM)	Activity (% conversion to product) [1]	Relative Activity
NH_4Cl	48.2	1
NH_4HCO_3	24.3	0.50
NH_4OAC	23.2	0.48

[1] Based on peak area of chorismate and new peak

Walsh et al. [114] proposed four possible merging mechanisms by which the reactions catalyzed by ADCS could occur: (1) addition of an enzyme nucleophile to the C6 position of chorismate in an S_N2' fashion followed by a second S_N2' displacement of the enzyme nucleophile, (2) addition of the enolpyruvyl side chain carboxylate to C6 to form a bicyclic lactone followed by a second S_N2' displacement, and (3) rigorous

S_N2'' displacement involving C2 addition and C4 substitution followed by a Mg^{2+} bound transition state. (4) a 1,5 SN_2'' addition of nucleophiles to C2 via a Mg^{2+}-bound transition state [135]. However, He et al. [136] demonstrated an active site nucleophile of K274 was employed by ADCS to add to C2 in a SN_2'' reaction and followed by a second SN_2'' displacement of the enzyme nucleophile at C2 by addition of NH_3 at C4 (Figure 2.9). This mechanism would explain why relatively poor inhibition of ADCS was observed for analogues of the transition state for C2/C4 concerted displacement, since the ADCS enzyme would interfere sterically with the inhibitor's C2 substituent. Another enzyme, ADCL, catalyzes the elimination of pyruvate from C3 to form PABA eventually (Figure 2.9).

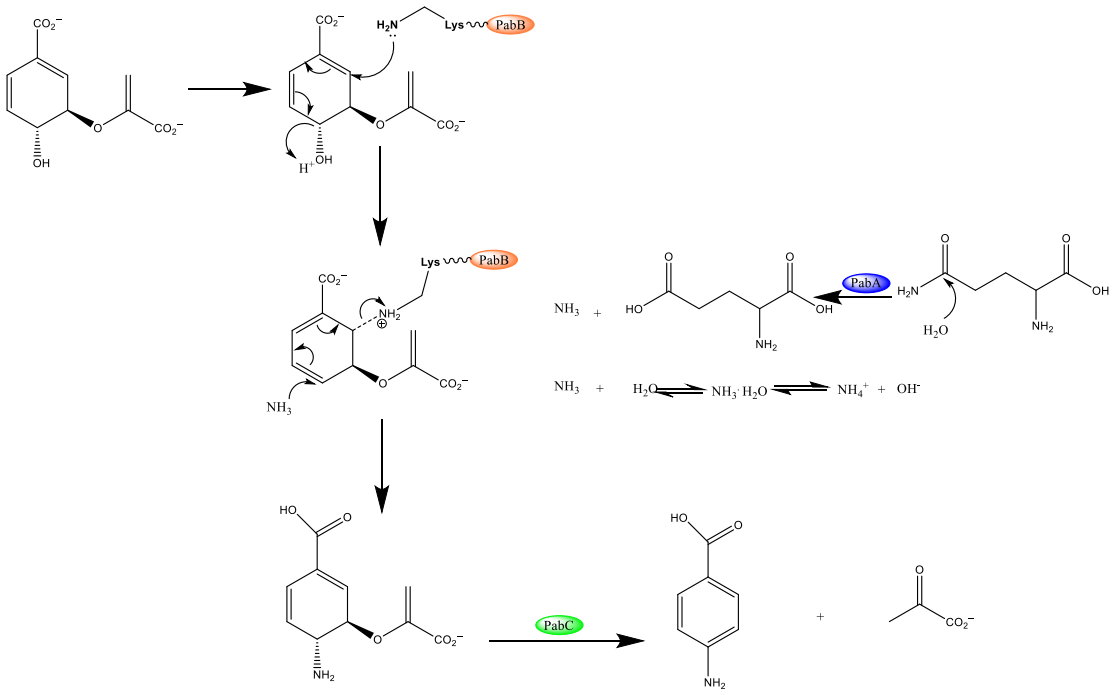

Figure 2.9 Mechanism of ADCS and ADCL.

2.4 Conclusion

Discovery and development of new anti-TB antibiotics are quite necessary. Capuramycins is a class of nucleoside antibiotics which show potent activity by inhibiting the bacterial MraY which involved in the biosynthesis of peptidoglycan cell wall. The first part of this dissertation describes the identification and characterization

of the initiation of biosynthetic machinery of abnormal radical polyamide ring for A-102395.

Resembling PabA and PabB bidomain architecture in *E. coli*, Cpr38 has been uncovered with same function of ADCS and structure characteristic for the biosynthesis of A-102395. A protein with identical bidomain architecture has been uncovered in the biosynthetic pathway of amicetin [137], pristinamycin [138], aureothin [139], candicidin/FR-008 [140], chloramphenicol [141], and albicidin [142]. Similarly, *cpr38* encoding a bidomain protein catalyzes a two-step reaction involving amidohydrolysis of L-Gln with ammonia incorporated into chorismic acid to generate ADC. Cpr12, carrying similarity to proteins annotated as ADCL or aminotransferase class IV of which ADC lyase belongs [120], catalyzes the elimination of pyruvate to form PABA. In summary, functional characterization of Cpr38 and Cpr12 indicated that they catalyze identical reactions as ADCS and ADCL, respectively. These data suggests a shared biosynthetic paradigm which forms PABA for biosynthesis of polyamide in capuramycin-type antibiotics and other NRPS or PKS pathway metabolite.

Chapter 3: Elucidating the Function and Mechanism of Cpr36 and Cpr37

3.1 Background

The unique machineries of both the ribosomal and the nonribosomal peptides make them ideal for their respective functions. The ribosomal peptide synthesis has several proofreading mechanisms that are absent in the nonribosomal system, which is crucial to the precision found in primary metabolism. Ribosomal peptide synthesis is normally restricted to a set of 20 aa as building blocks for proteins, whereas several hundred building blocks of NRPSs are known to date, which makes nonribosomal peptides carry structural diversity [88]. Some nonribosomally produced peptides contain unnatural amino acids and other molecules such as PABA not found in ribosomally produced peptides.

For the ribosomal production of proteins and peptides, the aa-tRNA synthetase first selects the cognate amino acid and then loads it onto the 2′ or 3′-hydroxyl group of the corresponding tRNA [143, 144]. Most aa-tRNA synthetases catalyze the first proofreading step in ribosomal protein synthesis with relaxed substrate selectivity. Because of this, they possess a site for hydrolysis of erroneously activated amino acids [145]. The ribosome subsequently selects the correct aa-tRNA with the help of the elongation factor (EF)-Tu during each cycle of polypeptide elongation [146]. After the ternary complex enters the acceptor (A)-site of the ribosome, the large ribosomal subunit stimulates GTP hydrolysis by EF-Tu as long as it detects complementary base-pairing between the mRNA and the cognate aa-tRNA. Consequently, GTP hydrolysis leads to a second proofreading step in ribosomal protein synthesis. Peptide bond formation occurs when the aa-tRNA has been docked to the A-site, regenerating the ribosome for following rounds of synthesis [147].

In the 1960s, scientists realized that the ribosomal machinery is not suitable to produce NRPS molecules. Gevers et al. conducted research that employed cell extracts of the producer strains and showed that these peptides were synthesized even in the presence

of RNases or inhibitors of the ribosomal machinery [148]. Producers of nonribosomal peptides are mostly found in gram-positive *Bacilli* and *Actinomycetes* genera, marine microorganisms or filamentous fungi [149]. Nonribosomal synthesis of peptides was thought to be restricted to these relatively simple organisms until recently, when people found that higher eukaryotes also harbor systems that bear a resemblance to NRPSs. For example, in mice, there is an aminoadipic acid semialdehyde dehydrogenase U26 which is a multifunctional enzyme comprising an A-domain, a PCP domain, and seven binding motifs for the redox factor pyrroloquinoline quinone [150].

NRPSs consist of an arrangement of modules, which are defined as sections of the NRPS's polypeptide chain that are responsible for the incorporation of one building block into polypeptide chain [151-153] (Figure 3.1). Since the amino acid chosen to be incorporated is determined by the modules which harbor all necessary catalytic functions, NRPSs are used simultaneously as a template and biosynthetic machinery. However, in bacteria modules, which consist of initiation and elongation domains, are usually distributed over several NRPSs whose genes are organized in a single operon [154]. Modules consist of initiation and elongation modules.

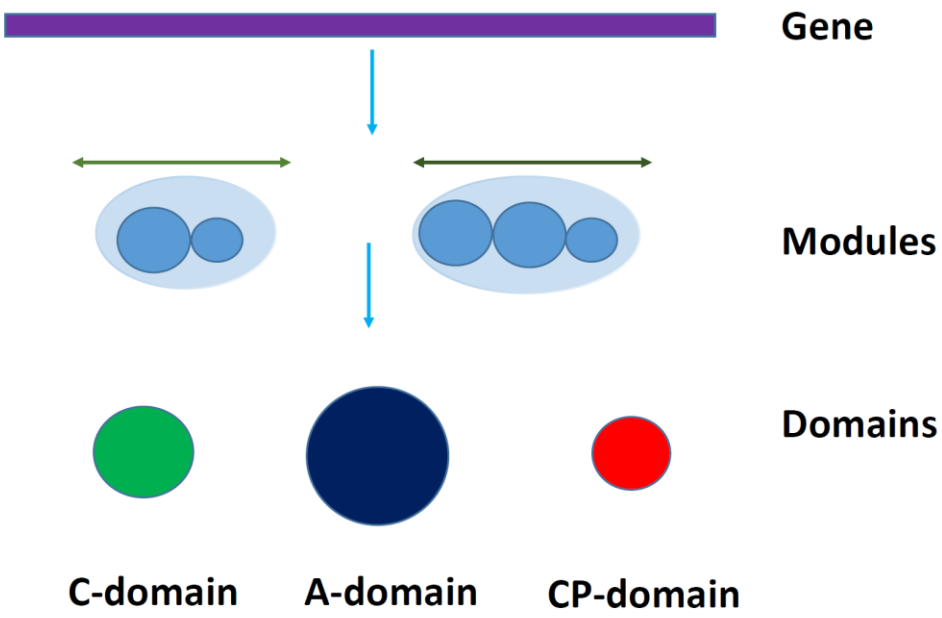

Figure 3.1 Modules are responsible for the incorporation of one amino acid. The modules can be subdivided into domains that harbor the catalytic activities for substrate activation (A-domain), covalent loading (CP-domain), and peptide bond formation (C-domain).

The enzymatic units that reside within a module are called domains (Figure 3.1). These domains which share some highly conserved sequence motifs catalyze at least three reaction steps: substrate activation, covalent binding, and peptide bond formation of nonribosomal peptide synthesis [155]. These “core-motifs” allow the identification of individual domains on the protein level [156]. One more exciting characteristic about domains is that they can be excised from the polypeptide chain and the resulting domains still retain enzymatic activity, which opens up prospects for the manipulation of the NRPS machinery.

Nonribosomal synthesis of peptides requires a minimum of the A-domain, C-domain and T-domain proteins. Much like the aa-tRNA synthetases, the A-domain selects the cognate amino acid and activates it as aminoacyl adenylate, however its structure is

unrelated to this class of enzymes [157, 158]. Reminiscent of the tRNA, the activated amino acid is then transferred to a PCP, which functions as the transport unit of the activated intermediate [159]. Normally, the C-domain, with an acceptor site for the nucleophile and a donor site for the electrophile, catalyzes the formation of the peptide bond [160, 161].

The A-domain, an approximately 550aa long sequence, is responsible for the selection of the amino acids that frame the product. A-domains activate amino or carboxy acid substrates by forming amino acyl adenylate while ATP is consumed [155, 157, 162]. Two crystal structures of A-domains have been resolved to date. Conti et al. tried to assign the amino acid residues that play a decisive role in the coordination of the substrate [163] by using the crystal structure of the phenylalanine-activating A-domain of the gramicidin S-synthetase A, GrsA from *Bacillus brevis.* Analysis of the residues that lie in a 100-aa stretch between cores A4, and A5 of the A-domain allows the prediction of A-domain selectivity by its primary sequence [158, 164]. Despite the similarities shared by the ribosomal and nonribosomal system during amino acid activation and enzymatic activity, the enzymes involved are unrelated in primary and 3D structures [88, 165-168]. Instead, A-domains are structurally related to the luciferase from photinus pyralis. Both enzymes comprise a large N-terminal and a small C-terminal subunit [88]. Stachelhaus demonstrated multimodular peptide synthetases on a polypeptide chain that act in concert concerning their order on this giant template. Novel peptide products can be produced by swapping single domains within the template [169].

The thiolation domain (T domain) of peptide synthetases, also known as the peptidyl carrier protein (PCP) domain, represents an integral part of these multi-enzyme domains and the site of 4′-PPT cofactor binding and substrate acylation [170-174]. This functional unit, consisting of about 80-100 amino acid residues, is bonded by aminoacyl substrates as carboxy thioesters. T domain is located in peptide synthetases directly downstream of the adenylation domains (A domains). In integrated peptide synthetases,

the activated amino acyl adenylate substrates on the A domain are transferred to the terminal cysteamine thiol group of the 4′-PPT and covalently attached to the side chain of the conserved serine residue [175]. There are a lot of studies demonstrating that the functional integrity of the domains is crucial for the multiple carrier models of nonribosomal synthesis [173, 176].

Further studies on posttranslational modification of T domain proteins via radiolabeled CoA led to the discovery of a superfamily of proteins that catalyze the conversion of apo forms to their holo ones [176]. Among this group of proteins are the gene products encoded by *gsp*, *sfp* and *bli*, which all are associated with bacterial operons encoding peptide synthetases which utilize CoA or its derivative as a common substrate and attain specificity through protein/protein interactions. Since transferases such as Sfp and Gsp are only associated with template-directed synthesis whereas are not essential proteins for the survival of the host, there was a prediction of the presence of homologous transferases that are specific for the modification of the essential ACP proteins of fatty acid synthesis and other proteins [177].

Also, PCPs exhibit a pI of ∼6.5, while ACPs are typically acidic proteins with a pI of ∼4, which is also reflected in the surface potential of the proteins: PCPs are nonpolar proteins with only a few charged residues on surface, whereas ACPs are negatively charged proteins [159]. Von et al. have demonstrated that PCPs show obvious differences depending on their localization, protein partners and accepted amino acids within the modular frame [178]. An unstructured C terminus that is possibly adjusted to different fatty acid substrates contributing to various structures was revealed by NMR studies on ACP from *Mycobacterium tuberculosis* [179]. The structure of ACP is concentration induced [180] and pH-dependent [181]. The protons of helix 2 exchange much faster than those in helix 1 or 3, which pave the way to a dynamic structural rearrangement [182]. The helix 2 of frenolicin ACP was also revealed to share such stunning instability [183]. As a result, CPs seem to be flexible proteins that can adjust to different substrates and proteins to contribute to the steering synthesis along the

NRPS machinery [184].

C-domain (~450 aa) is a classification of protein which are responsible for peptide bond formation between amino acyl substrates bound to PCPs of adjacent modules [185](Figure 4). Same as the adenylation and PCP domains in NRPS types of machineries, the C domain of peptide synthetases is an integral part of these multi-enzyme complexs [186]. The enzyme catalyzes the nucleophilic attack of the amino (or hydroxyl, imino) group of the activated amino acid bound to the downstream module with respect to the C-domain onto the acyl group of the amino acid tethered to the upstream T domain module. The C-domain possesses an acceptor site for the nucleophile and a donor site for the electrophile. This was confirmed by the crystal structure of the freestanding VibH which functions as an amide synthase that resembles NRPS C-domains [187].

Additionally, the accumulating sequence information of different peptide synthetase systems suggests that the C domain is responsible for the condensation of two individual amino acids activated on adjacent modules and elongation of the growing peptide chain [188, 189]. The C domain is restricted to the superfamily of peptide synthetases. Their distribution in multifunctional peptide synthetases seems to follow two simple rules: (I) Since in some scenario the two consecutive A domains are not located on the same enzyme, peptide bond formation has to occur between activated amino acids and the C domain is located at the N-terminus of the amino acid-accepting synthetase [190, 191]. (II) A C domain is always present between two adjacent activating units located on the same polypeptide [192-195]. Lagard et al. found that the C domains located at the N-terminus of accepting synthetases are less conserved than the internal ones [196]. Therefore, it is stimulating to speculate that the N-terminal C domains are also necessary for the accurate recognition of the preceding synthetase for protein/protein interaction; the observed sequence variations further confirmed such a specialized recognition reaction.

In addition to these essential domains, a number of tailoring enzymes act in the maturation and releasing of the NRPS-product.

3.2 Materials and Methods

3.2.1 Instrumentation, Chemicals and Bacterial Strains

PABA, anthranilic acid, lactate dehydrogenase (LDH), bicine, DTT, isopropyl-*β*-Dthiogalactopyranoside (IPTG), pyridoxal 5′-phosphate (PLP), CoA, Malonyl CoA, acetyl CoA, ATP, TCA, cysteamine, nucleoside bases, nucleosides, and nucleotides were purchased from Sigma(St. Louis, MO). [^{3}H]PABA, buffers, salts, and media components were purchased from Fisher Scientific. Synthetic oligonucleotides were purchased from Integrated DNA Technologies. Amino acids reagents were purchased from Sigma Aldrich (St. Louis, MO). All other reagents used were analytical grade. Restriction enzymes, DNA ligase, and Vent DNA polymerase were from New England Biolab. Wizard® Plus SV Minipreps DNA Purification Systems, Wizard® SV Gel, and PCR Clean-Up System were purchased from Promega (Madison, WI, USA). The protein native gel Novex™ Value™ 4-20% Tris-Glycine Mini Gelswas was purchased from Thermo Fisher Scientific Co., Ltd. (Waltham, MA). Cayman malachite green phosphate assay kit was purchased from Cayman Chemical Co., Ltd. (Ann Arbor, MI). DNA sequencing was performed and analyzed by High Throughput DNA Sequencing service was provided by ACGT Inc. (Wheeling, IL). UV/Vis spectroscopy was recorded on an Ultrospec 8000 spectrometer (GE, Pittsburgh, PA, USA) or performed with a Bio-Tek μQuantmicroplate reader using Microtest™ 96-well plates. HPLC for protein analysis was performed with one of two systems: a Waters Alliance 2695 separation module (Milford, MA) equipped with a Waters 2998 diode array detector and an analytical Apollo C4 column (250 x 4.6 mm, 5 μm) or Dionex Ultimate 3000 HPLC system equipped with a diode array detector Apollo C4 column (250 x 4.6 mm, 5 μm) purchased from Grace (Deerfield, IL). Semipreparative HPLC was performed with a Waters 600 controller and pump (Milford, MA) equipped with a 996

diode array detector, 717plus autosampler, and an Apollo C-18 column (250 × 10 mm, 5 μm) purchased from Grace (Deerfield, IL). HR-ESI-MS spectra were recorded on an AB SCIEX Triple TOF 5600 System (AB Sciex, Framingham, MA, USA). LC-MS was conducted with an Agilent 6120 Quadrupole MSD mass spectrometer (Agilent Technologies, Santa Clara, CA) equipped with an Agilent 1200 Series Quaternary LC system and an Eclipse XDB-C18 column (150 × 4.6 mm, 5 μm). All solvents used were of ACS grade and purchased from Fisher Scientific (Pittsburgh, PA) or Pharmco-AAPER (Brookfield, CT). TLC silica gel plates (60 F254) were purchased from EMD Chemicals Inc. (Darmstadt, Germany).

3.2.2 DNA Extraction, Genome Sequencing, and Analysis

To ensure the accuracy of the gene of *cpr36* and *cpr37*, the whole genome of A-102395 was extracted by the following procedures: SANK 60206 was grown without any antibiotics on ISP4 agar for two weeks at 28 °C to allow sporulation. The spores were collected from ISP4 agar and added into 5 mL 2×YT broth with overnight culture. 1 mL of an overnight culture was added into a 1.5 mL microcentrifuge tube. The tube was centrifuged at 16,000×g for 2 minutes to pellet the cells and remove the supernatant. The cells were resuspend thoroughly in 480 μL of 50mM EDTA and the lytic enzyme was added to the resuspended cell pellet. Then, a mixture of 60 μL of 10mg/mL lysozyme and 60 μL of 10mg/mL lysostaphin was added to increase the lysis efficiency. The sample was incubated at 37°C for 60 minutes and centrifuged at 16,000×g for 2 min. The supernatants were removed and 600 μL of Nuclei Lysis Solution was added. The samples were resuspended. After incubating at 80°C for 5 minutes, the cells were lysed and then cooled to room temperature for about 20 minutes. 3 μL of RNase Solution was added to the cell lysate. The tube was inverted the tube 2–5 times to mix. The cell lysate was incubated at 37°C for 60 minutes and cooled to room temperature. 200 μL of Protein Precipitation Solution was added to the RNase treated cell lysate. Then the sample was vortexed vigorously at high speed for 20 seconds to mix the

Protein Precipitation Solution with the cell lysate. The sample was then incubated on ice for 5 minutes. The cell lysate was centrifuge at 16,000×g for 5 minutes and the supernatant was transferred into a clean 1.5 mL microcentrifuge tube containing 600 µL of room temperature isopropanol. The tube was gently inverted until the thread-like strands of DNA form a visible mass and centrifuged at 16,000×g for 2 minutes. The supernatant was carefully pour off and the tube was drained on clean absorbent paper. 600 µL of room temperature 70% ethanol was added and the tube was gently inverted several times to wash the DNA pellet gently. Then the tube was centrifuged at 16,000×g for 2 minutes.100 µL of DNA Rehydration Solution was added to the tube to rehydrate the DNA by incubating at 65°C for 1 hour. The solution was periodically mixed by gently tapping the tube. Finally, the DNA was stored at 2–8°C.

3.2.3 Cloning, Overexpression and Purification of Proteins of Cpr36 and Cpr37

The genes, *cpr36,* was amplified by PCR from genomic DNA extracted from SANK 60206 using Herculase II Fusion DNA Polymerases from Agilent Genomic with supplied buffer , 200 µM dNTPs, 5% DMSO, 10 ng template, 10 U DNA polymerase, and 200 nM each of the primer pairs in table 3.

Table 3.1 Primers for amplification of *cpr36* and *cpr37*.

Name	Sequence
Cpr36_forward	5′-GGTATTGAGGGTCGCATGATCACTGTTGAGGATGT CAAG-3′
Cpr36_reverse	5′-AGAGGAGAGTTAGAGCCTCACCTCTCCACCGAGA CAC-3′
Cpr37_forward	5′-GATACGATTAATATGGCCGCC GGTAGTTC-3′
Cpr37_reverse	5′-TAGAGTGAATTCTCAATGATCAGCGTGCGC-3′

Purified PCR products were ligated to pET30 Xa/LIC (Novagen) following the standard protocol and confirmed by DNA sequencing (ACGT, INC). The PCR program begins at 94 °C for 2 min, followed by 10 cycles of 94 °C for 45 s, 50°C for 45 s and 72°C for 90 s, and then 15 cycles of 94 °C for 45 s, 55°C for 45 s and 72°C for 90 s, and a final extension at 72°C for 5 min. The resulting linear PCR products were purified by electrophoresis on 0.5% agarose gel, and the desired DNA fragment was extracted following the above-mentioned manufacture's standard protocol. Plasmids pET30-*Cpr36* was introduced into *E. coli* BL21 (DE3) cells and the transformed strains were grown after picking one single colony into a 5 mL tube filled with LB supplemented with 50 μg/mL kanamycin overnight. Following inoculation of 500 mL of LB with 50 μg/mL kanamycin, the cultures were grown at 37 °C until the cell density reached an OD600 of ~ 0.5 when expression was induced with 0.25mM IPTG. Cells were harvested after an overnight incubation at 18 °C and lysed in lysed buffer(100 mM KH_2PO_4, 300 mM NaCl, 10 mM imidazole, pH 8.3) using a Qsonica sonicator (Qsonica LLC, Newtown, CT) for sonication for a total of 10 min at 40% amplitude with 30 s pulses separated by 60 s rest periods. Following centrifugation, the protein was purified using affinity chromatography with HisPurTM Ni-NTA agarose (Thermo Scientific,

Rockford, IL), and proteins were firstly washed with Wash Buffer (100 mM KH_2PO_4, 300 mM NaCl, 20mM imidazole, pH 8.3) and eluted with Elution Buffer(100 mM KH_2PO_4, 300 mM NaCl, 200mM imidazole, pH 8.3). Purified proteins were concentrated and then buffer exchanged into desalting buffer (25 mM KH2PO4, 100 mM NaCl, pH 8.3) using Amicon Ultra 10000 MWCO centrifugal filter (Millipore) and stored with glycerol (40%) at -20 °C. Protein purity was analyzed as by 12% acrylamide SDS-PAGE; His6-tagged proteins were utilized without further modifications. *Streptomyces* verticillus phosphopantetheinyl transferases (Svp) was purified similarly except for the selective antibiotic was 50 μg/mL ampicillin. Svp enzyme was purified by following the same protocol for *cpr36*.

The gene for *cpr37* was amplified from genomic DNA extracted from *Streptomyces* sp. SANK 60206 above-mentioned by PCR using Herculase II Fusion DNA Polymerases. The PCR program included an initial hold at 95 °C for 3 minutes, followed by 25 cycles of 95°C for 30 s, 60°C for 30 s, and 72 °C for 60 s and a final hold at 72 °C for 5 minutes.

The gel-purified 1.4kb PCR product was digested with AseI-EcoRI and the 1.4-kB DNA fragment purified and ligated to the identical sites of pXY200 plasmid. The resulting combinant plasmid was transformed into Nova Blue cells and spread onto an LB plate with 50 μg/mL Apramycin. After 4 days, positive transformants were confirmed by colony PCR using InstaGene Matrix from Bio-Rad (Hercules, CA) and LA-Taq polymerase with GC buffer II from Takara Bio Inc. (Shiga, Japan) and ATGC DNA sequencing service (Wheeling, IL, US). After plasmid DNA extraction, pXY200 plasmid was transformed (with *cpr37* gene to be expressed) to *Streptomyces* lividans TK24 protoplasts using PEG-mediated protoplast transformation and plated on R2YE. Several colonies' spores were picked to do colony PCR and gene sequencing for fidelity confirmation. After 6 days at 28 °C, positive transformants were confirmed by colony PCR using InstaGene Matrix from Bio-Rad (Hercules, CA) and LA-Taq polymerase with GC buffer I.

A single colony was utilized to inoculate 50 mL R2YE containing 50 μg/mL apramycin, grown for 3 days at 28 °C at 250 rpm, and 2 mL transferred to fresh 150 mL containing 50 μg/mL apramycin. Following growth for 3 days at 28 °C at 250 rpm, protein expression was induced by addition of thiostrepton (25 μg/mL) and the culture was incubated for another 48 h before harvesting. The cells from 400 mL of culture were collected by centrifugation and washed with 10% glycerol before lysis. The pellet was thoroughly resuspended in ice-cold lysis buffer(100 Mm Tris-HCl, 300 mM NaCl, 10 mM imidazole, pH 8.3) supplemented with 10 mg/mL of lysozyme added to the suspension. One thing to note here is that any buffer that contains phosphate is excluded to get rid of phosphate contamination after protein purification. After incubation at 30 °C for 30 min, the cell suspension was mixed by pipetting and lysed using a Qsonica sonicator (Qsonica LLC, Newtown, CT) for a total of 20 min at 40% amplitude with 30 s pulses separated by 30 s rest periods. The protein purification and storage was as described above except that the phosphate buffers above were substituted with wash buffer (100 mM Tris-HCl, 300 mM NaCl, 20mM imidazole, pH 8.3), elution buffer(100 mM Tris-HCl, 300 mM NaCl, 200mM imidazole, pH 8.3) and the desalting buffer (25 mM Tris-HCl, 100 mM NaCl, pH 8.3).

3.2.4 Reaction of Transformation from Apo-Cpr36 to Holo-Cpr36

Reactions consisted of 10 mM Tris-HCl (pH 7.5), 5 mM $MgCl_2$, 250μM Cpr36, 150 μM Svp, and 0.5mM CoA, with the water up to the volume of 100μL at 30 °C for one hour.

3.2.5 HPLC Analysis of Transformation from Apo-Cpr36 to Holo-Cpr36

Following centrifugation to remove any participation on the bottom which may block the HPLC column, the reaction was first monitored and analyzed by using reverse-phase HPLC equipped with an analytical HyPURITY™ C4 column(150 mm × 4.6mm, Fisher Scientific Co., Ltd.). The elution of proteins was monitored based on their

absorbance at 220 nm. Following solvents were prepared: HPLC Solvent A: Filter 1 L ddH_2O with 0.5 µM filter. Add trifluoroacetic acid (TFA) to a final concentration of 0.1% (vol/vol). HPLC Solvent B: Prepare 1L of HPLC-grade acetonitrile (ACN), 0.1% TFA (vol/vol).

Samples to be analyzed were prepared. These may include samples prepared as described above for the in vitro phosphopantetheinylation of Cpr36. The enzyme concentrations were 250µM for Cpr36 and 150 µM for Svp. 100 µL of sample was injected onto the column. HPLC was performed according to using a C-4 reversed-phase column (Grace, Deerfield, IL) under the following conditions: Firstly the column was equilibrated first with a mixture of 80% solvent A/ 20% Solvent B at a flow rate of 1 mL/min. Then a series of linear gradients in the following manner (beginning time and ending time with linear increase to % B): 0-5 min, 20% B; 5-20 min, 80% B; 20-30min, 80% B; 30-41 min, 20% B;

The absorbance at 220 nm was monitored, with peaks corresponding to the elution of proteins. The peak corresponding to the carrier proteins of interest were collected in colorless microcentrifuge tubes to minimize the potential for 'extractables' by ACN, flash frozen in a CO_2 (s)/ethanol bath, and stored at −80°C for further analysis (see below for ESI-TOF MS analysis procedure). Holo form Cpr36 peaks were identified by comparison to apo form. The conversion from the apo to holo form was determined by calculating the area under the peaks.

3.2.6 Protein Native gel analysis of Transformation from Apo-Cpr36 to Holo-Cpr36

The protein native gel electrophoresis analysis was conducted according to the manufacturer's instructions at room temperature. The concentration of the protein in the reaction system was determined by following the protocol of the Quick Start™ Bradford Protein Assay from Bio-Rad (Hercules, CA). 1X Novex™ Tris-Glycine SDS Running Buffer was poured into the electrophoresis tank. Samples contained glycerol

and bromphenol blue and the pH was adjusted to between 6.0 and 7.5. Voltage was set at 90mV until the tracking dye reached the gel bottom. Gels were stained overnight in 50% methanol, 10% acetic acid, 0.1% Coomassie Blue and 40% H_2O, and destained in 10% methanol, 20% acetic acid, 70% ddH_2O for a minimum of 24 hrs. Theoretically, ACP is an acidic protein and stains poorly using most protocols. Several combinations of methanol, acetic acid, and different stains were tried, and the above procedure was determined empirically to be the most satisfactory.

3.2.7 Protein Mass Spectrometry Analysis of Transformation from Apo-Cpr36 to Holo-Cpr36

Following the protein native gel analysis, individual peaks were collected from HPLC and for the further high-resolution mass spectrum characterization. HPLC-purified Cpr36 were kept at −80°C for analyzing. Proteo-Mass™ Protein ESI-MS Calibration Kit was purchased from Sigma-Aldrich which includes ACN, sinapinic acid, 0.1% TFA and protein standards. 5mL 50% ACN/ 0.05% TFA was prepared in a colorless microcentrifuge tube. 5mL 10 mg/mL sinapinic acid matrix solution was mixed with 600 μL 50% ACN/ 0.05% TFA solution. The mixture was vortexed vigorously for 1 min. The mixture was stored in the dark after settling for 10 min at the bench at 22°C. Prior to ESI-TOF MS analysis, the ACP-containing microcentrifuge tubes were removed from −80°C, opened, and placed inside 50 mL polypropylene tubes and a Kimwipe® was placed over the top of the polypropylene tube and secured with a rubber band. Resulting tubes were then placed inside a lyophilizing vessel and lyophilized until the ACN has evaporated (~24 h). The lyophilized protein was resuspended in 20 μL of ddH_2O and placed on ice until further use. Protein standards cytochrome C and bovine insulin were prepared by following the manufacturer's instructions.

The samples for analysis were prepared by adding 5 μL of a sample to a colorless microcentrifuge tube containing 5 μL of sinapinic acid. Matrix and sample were mixed by pipetting, and 2 μL of each sample was spotted onto the sample plate. The plate was

allowed to dry completely before being inserted into the instrument chamber. The manufacturer's instructions were followed for using the mass spectrometer, with the instrument set to linear, positive-ion mode. Specifically, ESI mass spectra were recorded on a Thermo Scientific™ Q Exactive™ Orbitrap Mass Spectrometer. Resolution was set at 100,000 (at 400 *m/z*). Samples were introduced through direct infusion using a syringe pump with a flow rate of 5μL/min. The instrument was calibrate by using the standards.

3.2.8 PPi Exchanging Assay to Study the Enzyme Activity of Cpr37

The ATP-[^{32}P]pyrophosphate (PPi) exchange assays mixture for the reaction consisted of 75mM Tris-HCl (pH 8), 5 mM $MgCl_2$, 2 mM ATP, 0.2 mM [^{32}P]PPi (54.39 Ci/mmol), containing 1 μCi [^{32}P]PPi, 1 mM EDTA, 1 mM carboxylic acid substrate or substrate analog and 100 nM enzyme of Cpr37. The reaction was performed following the described procedures. Following the initiation by the addition of enzyme of Cpr37, reactions were allowed to proceed for 20 min at 30 °C, and then quenched with charcoal suspensions 500 μL of 1.6% (w/v) activated charcoal, 4.5% (w/v) tetrasodium pyrophosphate, and 3.5% perchloric acid in water. The charcoal was precipitated by centrifugation, washed twice with quenching buffer with 40 mM sodium pyrophosphate in 1.4% (v/v) perchloric acid (100 mL), water (200 mL), and ethanol (200 mL). Filter papers were added to 10 mL scintillation fluid and radioactivity was quantified and subjected to liquid scintillation counting. The amount of bound radioactivity was converted to reaction rate using the specific radioactivity of the [^{32}P]pyrophosphate and the length of reaction time. Data represent 3 duplicates with an S.D. of <10%.

3.2.9 Inorganic Pyrophosphatase Combined with Malachite Green Phosphate Assay

To determine the real time specific concentration of free phosphate which is vital for this kinetic study, a malachite green phosphate assay method was followed and improved [197]. Figure 3.2 shows the specific experimental procedure. The reaction

mixture of enzyme of Cpr37 consisted of 75mM Tris-HCl (pH 8), 5 mM $MgCl_2$, 0.5mM ATP, 1 mM EDTA, 0.5mM PABA, and 50 nM enzyme of Cpr37, 0.2U of pyrophosphatase was purchased from New England Biolabs (Ipswich, MA) and water was added up to a volume of 100 μL. To determine the standard curve, 1 M malachite green phosphate standard was diluted at the ratio of 1:100 by adding 10 μL to 990 μL assay buffer (75mM Tris-HCl, 5 mM $MgCl_2$, pH 8). The concentration of this solution was10 mM and stored at 4°C for just one day use only. Eight clean test tubes were numbered as D2 and D3 (dilution tubes) and the others are marked as S1 through S6 (standard tubes). 990 μL of assay buffer was added into tube D2 and mixed. 500 μL of D2 was transferred into D3 and mix thoroughly and the phosphate concentration was decreased to 50 μM. The standards were serially diluted by removing 500 μL from tube D3 to tube S1. Next, the rest of the tubes were done in the same manner for tubes S2-S5 while leaving S6 as the blank. The six-point standard curve was prepared based on the absorbance value from the microplate reader set to 620 nm. The standard curve solutions had the same final volume as the samples. 5 μL of malachite green acidic solution was added to plate well and tapped gently to mix thoroughly. The plate was incubated for 10 minutes at room temperature and 15 μL of malachite green blue solution was added to each well and mixed by gently tapping. Then the plate was incubated for 20 minutes at room temperature. The absorbance was determined separately in a spectrophotometer with plastic curettes of each well using a microplate reader set to 620 nm. The assay was done in triplicate each time and set up negative control reactions (without Cpr37) were set up in the same buffer as the experimental samples.

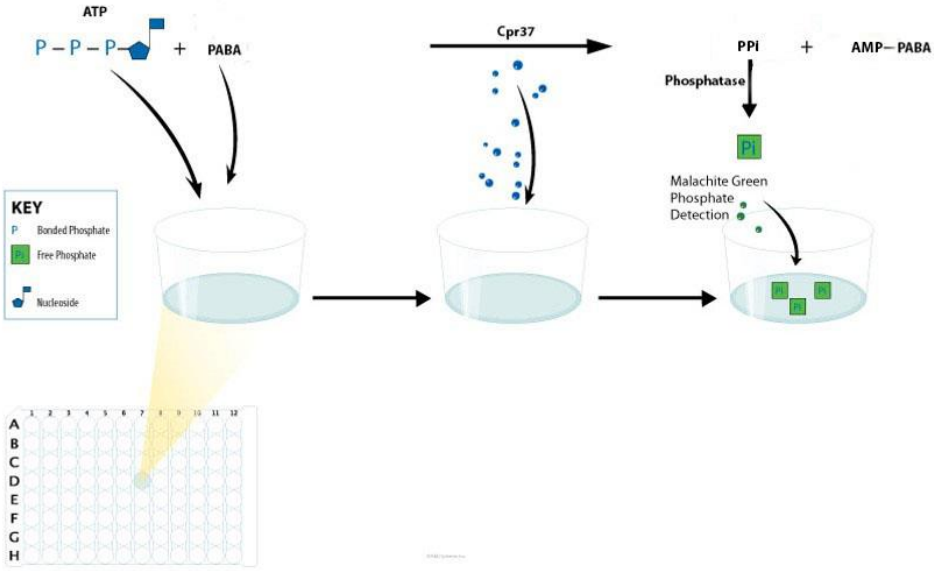

Figure 3.2 Illustration of free inorganic pyrophosphate malachite green phosphate assay [198].

3.2.10 Determination of the Incorporation of PABA onto the Holo-Cpr36

To determine whether PABA was incorporated onto the Holo-Cpr36. Here, we applied the following protocols: reaction mixtures (100µL), which contained 50 mM Tris-HCl (pH 8), 5 mM $MgCl_2$, 100 µM ATP, 100 µM CoA, 10µM Cpr36, 5 µM enzyme of Svp and water up to 100 µL was incubated at room temperature for 30 min until the apo-form of Cpr36 was converted to holo form. 4 µM [^{3}H]PABA (26 Ci/mmol) was added into the reaction system. Reactions were terminated after 0, 10, 20, 30, and 40 min by adding 100µL TCA into the reaction. The precipitate was washed twice with 200 µL washing buffer 40 mM sodium pyrophosphate in 1.4% (v/v) perchloric acid (100 mL), water (200 mL), and ethanol (200 mL) and washing buffer was sucked out. Filter papers were added to 5 mL scintillation fluid and radioactivity quantified as above mentioned. Data represented 4 duplicates with an S.D. of <10% before processing and liquid scintillation counting (PerkinElmer, Waltham, MA).

3.2.11 Cysteamine-Promoted Cleavage of Cpr36-Bound Polyketides

A method was followed to cleave the small molecule from the Cpr36-PABA bounded polyketide as previously described [199]. 1M solution of cysteamine-HCl (Fluka/ Sigma-Aldrich, St. Louis, MO) was prepared in nickel resin elution buffer at neutral pH. Freshly prepared solutions of reaction pools of Cpr36 in nickel elution buffer were supplemented with the buffered cysteamine solution to a final concentration of 0.2 M cysteamine, capped, and wrapped in aluminum foil. Cysteamine treatment proceeded for 20 h at 4 °C. Samples were then extracted twice with ethyl acetate and concentrated by rotary evaporator. The residue was immediately resuspended in an appropriate solvent, filtered, and analyzed by HPLC/LC-MS. HPLC was performed according to using a C-18 reversed-phase column under the following conditions: a series of linear gradients from 0.1% Formic acid in 100% water (A) to 0.1% Formic acid in acetonitrile (B) in the following manner (beginning time and ending time with linear increase to % B): 0-16 min, 5%-100% B; 16-20 min, 100% B; 20-30 min, 100%-5% B. The flow rate was kept constant at 0.4 ml/min, and elution was monitored at 254 nm.

3.3 Results and discussion

In an traditional NRPS pathway, the chemistry logic is undergoes as following: A domains catalyze a two-step, ATP-dependent reaction that involves the activation of the carboxylate group of the amino acid substrate to form an aminoacyl-AMP intermediate. The amino acid is transferred to the 4′-PPT arm of the neighboring peptidyl carrier protein (PCP) domain (Figure 3.3). The previous study in the Van Lanen's group demonstrated that ~60% sequence similarity of Cpr36 with AcmD which functions as an aryl carrier protein. In this thesis, we proposed that the enzyme of Cpr37 functions as adenylation domain protein. It initiates the PABA and ATP to form a PABA-AMP intermediate. Meanwhile, the carrier protein is catalyzed by Svp enzyme to add the 4′-PPT of the CoA to its serine residue. Then the PABA-AMP is transferred and incorporated into the carrier protein of Cpr36 with AMP-forming as the side product

(Figure 3.3). The *cpr36* and *svp* gene was then cloned and expressed in *E. coli* to yield the purified protein with the expected size (Figure 3.4, His-tag included).

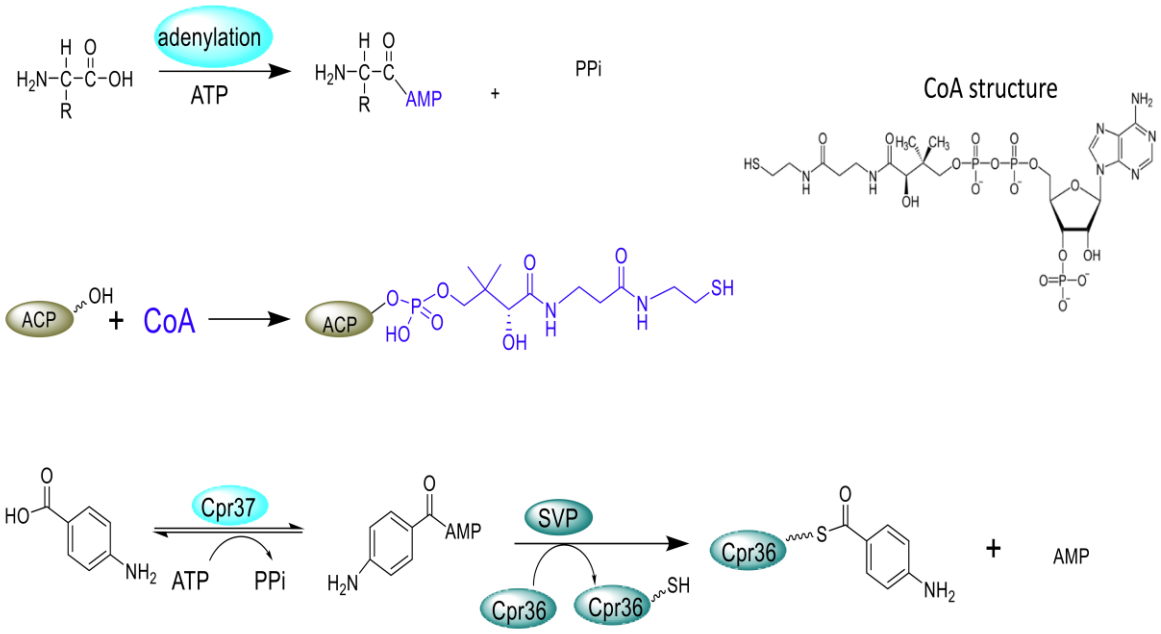

Figure 3.3 Proposed reaction mechanism for Cpr37 and Cpr36.

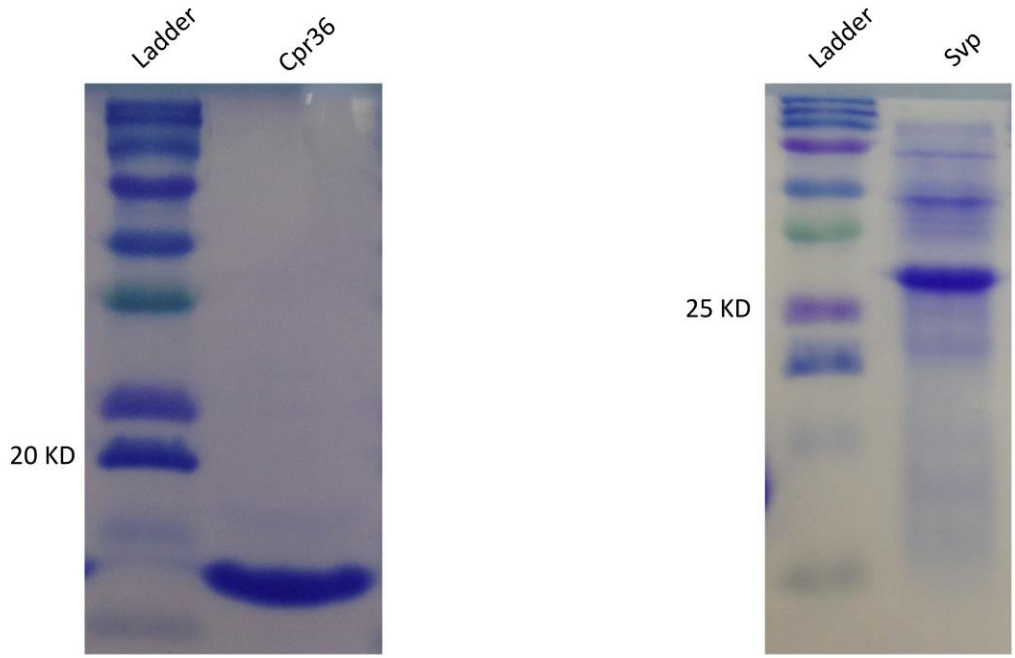

Figure 3.4 SDS-PAGE analysis of purified protein of Cpr36 and Svp (expected MW: Cpr36, 15 kD; Svp, 30 kD).

To investigate the function of Cpr36, HPLC was utilized for the analysis. In the reaction mixture adding Svp enzyme, Cpr36 revealed a distinguished chromatogram peak at the retention time of 17.8 min whereas untreated Cpr36 eluted at 16.4 min, suggesting that Cpr36 was post-translationally activated with 4′-PPT with the formation of the holo form of Cpr36 which may have different confirmations. Further assays of HPLC/ESI MS–analyzed enzyme reactions using a C4 reversed-phase column provided more insights into the reaction pathway (Figure 3.6). In untreated Cpr36, one main chromatogram peak was detected, with an observed molecular masses of 14,388 Da at the retention time of 17.8 min (Figure 3.6A). However, in the reaction of Cpr36 treated with Svp, one chromatogram peak was observed with the molecular weight of 14,729 Da at the retention time of 16.4 min (Figure 3.6B). The difference between these masses (341 Da) was in good agreement with the molecular mass of 4′-PPT (340 Da). This demonstrated that the chromatogram in Figure 3.5 (I) corresponded to an apo form of Cpr36 and that apo-Cpr36 was post-translationally activated to form the holo form of

Cpr36 catalyzed by the Svp enzyme (Figure 3.5II). Also, in untreated Cpr36, one small chromatogram peak of 14,729 Da was also detected, suggesting that there exists a small portion of activated holo form of the carrier protein purified from bacterial cells.

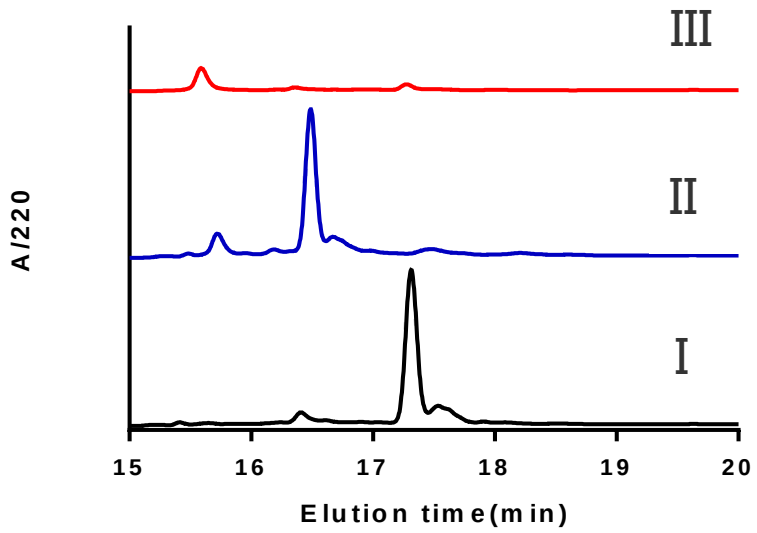

Figure 3.5 Transformation between apo- and holo-form of Cpr36. HPLC analysis of transformation of mutated Cpr36. I, apo-ACP as purified from *E. coli.*; II, mutated form of Cpr36 resulted from Svp catalyzing and adding the 4′-PPT arm from CoA into the apo mutated Cpr36; III, Svp as purified from *E. coli*. HPLC/ ESI MS–analyzed enzyme reactions using an analytical HyPURITY™ C4 HPLC column (150 mm × 4.6mm, Fisher Scientific Co., Ltd.) reversed-phase at 25 °C, at a flow rate of 0.4 mL min^{-1}.

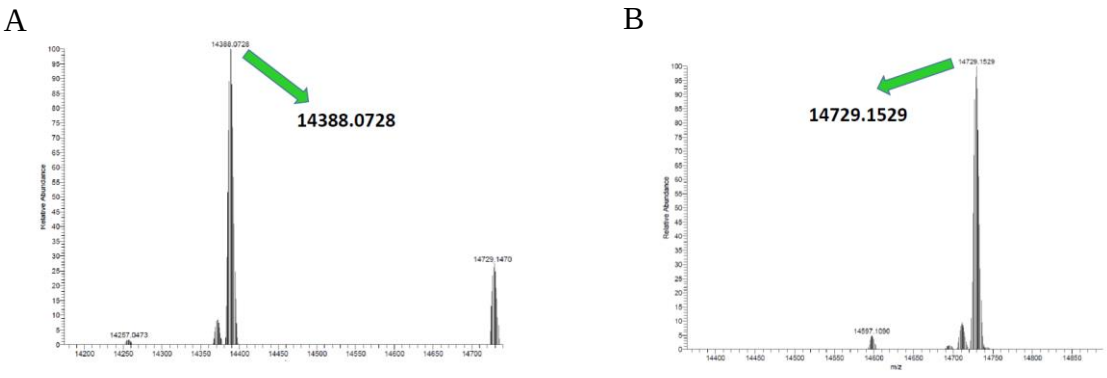

Figure 3.6 Molecular mass change of Cpr36 after the reactions. (A) Apo form of Cpr36. (B) Apo form of Cpr36 treated with Svp and CoA.

Although the post-translationally activation of Cpr36 and adduction of a 4′-PPT was confirmed by the HPLC and ESI MS data, there was an unexpected molecular mass discrepancy of 32 Da between the observed molecular weight of Cpr36 and the theoretical one. We assume that this may be due to a gene mutation during the process of gene overexpressing or extended storage time as sequencing had been conducted to confirm accuracy after gene cloning. Mutations are not abnormal when conducting gene cloning and genomic DNA storage. The likely source of the mutation could also be the polymerase or other steps while handling the insert. We just found that the error rate will increase with excessive Mg^{2+} and high concentrations of dNTPs. Additionally, the gel-purifying of the insert and the time spent illuminating the insert with UV light are likely to introduce mutations. The assumption of a gene mutation was confirmed by the sequence alignment of mutated *cpr36* to the gene sequence of *cpr36* (ID:AKC92648.1) from NCBI using Clustal Omega, which revealed that there is a mutation from Met 63 to Val 63(Figure 3.7). This mutation would lead to the molecular weight loss of 32 Da, which perfectly explains the 32 Da discrepancy.

Figure 3.7 Sequence alignment of mutated *cpr36* to gene sequence of *cpr36* (ID:AKC92648.1) from NCBI using Clustal Omega. The mutation sites were highlighted in red which lead to an amino acid switch from methionine to valine.

Although gene sequencing data revealed that there is a mutation from the hydrophobic amino acid of Met to the hydrophobic amino acid of Val, it seems that a mutation hadn't influenced post-translationally activation of Cpr36 by adduction of 4′-PPT. This result further strengthened the statement that 4′-PPT transferase activated the T domain only by the addition of the 4′-PPT on a conserved serine residue rather than some other residues [77].

To get the wild type Cpr36 without any mutation, we optimized all the experiment conditions (including remaking new cosmids and using new high fidelity DNA polymerase). We obtained the PCR DNA product and some soluble protein with the

expected size (Figure 3.8). The authenticity of the gene was double confirmed by gene sequencing and sequence alignment of *cpr36* (ID:AKC92648.1) from NCBI to the wild type Cpr36 we overexpressed and purified using Clustal Omega (Figure 3.9).

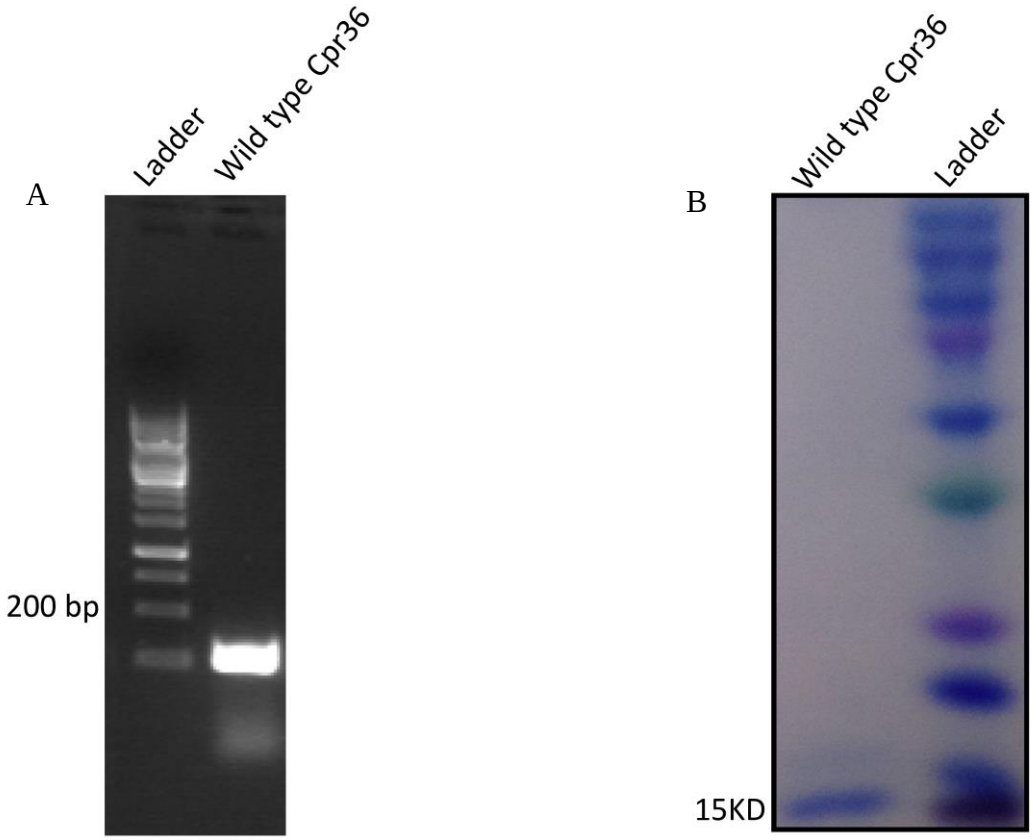

Figure 3.8 Gene cloning and overexpression of wild type *cpr36*. (A) 1% agarose gel electrophoresis analysis of the PCR product of wild type *cpr36* (expected length 256 bp). (B) SDS-PAGE analysis of the purified protein of wild type Cpr36 (expected MW of 15 kD).

Figure 3.9 Sequence alignment of Cpr36 (ID:AKC92648.1) from NCBI to wild type-Cpr36 using Clustal Omega.

Unsurprisingly, Figure 3.10 shows HPLC traces for apo- (I) and holo-wild type(II) Cpr36. The addition of 4′-PPT prosthetic groups to the apo-wild type Cpr36 results in a shorter elution time for holo-wild type(II) Cpr36 which of around 16 min. The percent conversion from the apo- to holo- forms determined by calculating the area under the peaks revealed that almost 100% of the apo form of Cpr36 was converted to the holo form.

The characterization of apo-wild type Cpr36 was further confirmed by using electrospray ionization protein mass spectrometry (ESI-MS) (Figure 3.11A) (observed,

14420Da; predicted, 14420 Da). The quantitative conversion of the pure apo-wild type Cpr36 to the holo form was validated by the ESI protein mass spectrum with the expected molecular weight gain of 341 Da of the 4′-PPT arm (observed,14761 Da; predicted,14760 Da) (Figure 3.11B).

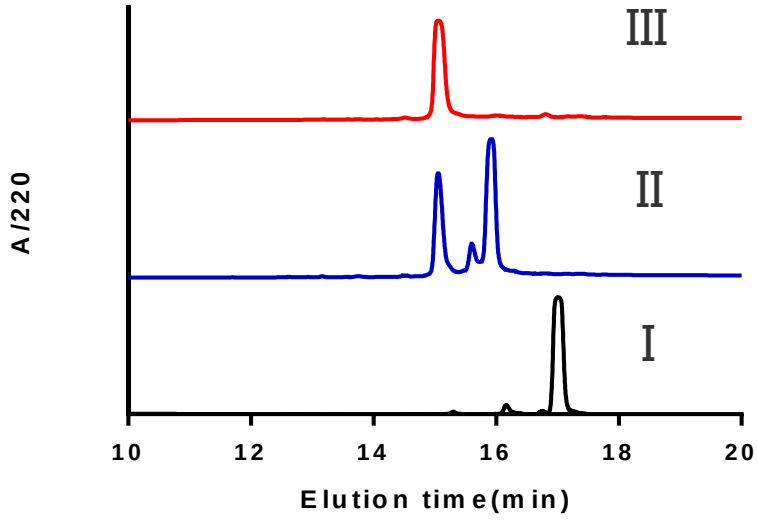

Figure 3.10 HPLC analysis of transformation of wild type Cpr36. I, apo wild type Cpr36 as purified from *E. coli*; II, holo form of wild type Cpr36 resulted from Svp catalyzing by adding 4′-PPT arm from CoA into the apo wild type Cpr36; III, Svp as purified from *E. coli*.

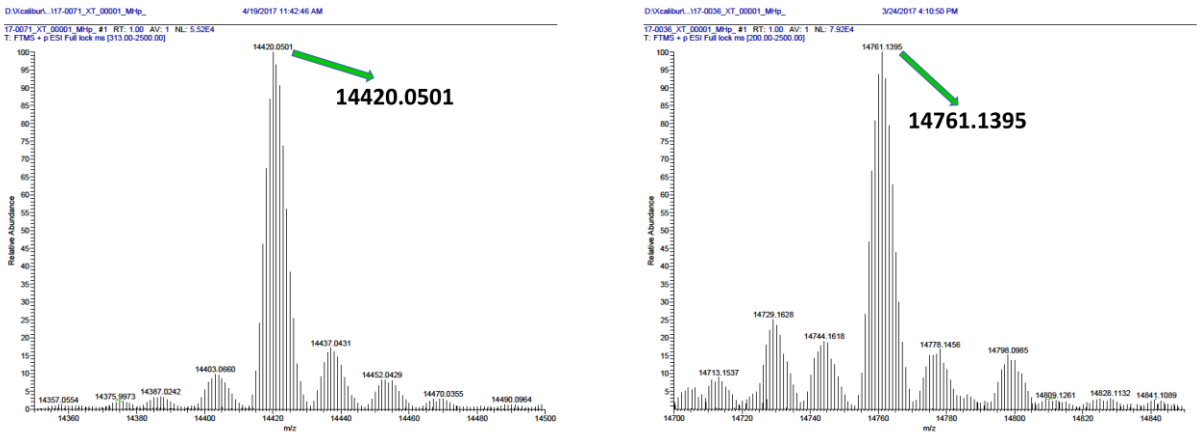

Figure 3.11 Molecular mass change of wild type Cpr36 after the reactions. (A) Apo form of wild type Cpr36. (B) Apo form of wild type Cpr36 treated with Svp and CoA.

Additionally, protein native polyacrylamide gel electrophoresis was applied for the analysis of the transformation of Cpr36 (Figure 3.12). The apo form of Cpr36 (lane1&4) showed distinguished mobility comparing to the holo form (lane2&3). The holo form of Cpr36 showed faster mobility, which hinted that holo Cpr36 may have a more compact structure compared to the apo one. This is consistent with the previously reported data about the acyl carrier protein [200]. Interestingly, we also observed a band which showed the same mobility comparing to the apo form of Cpr36, which provided more evidence that there is also small portion of holo form of carrier protein in the bacterial cell. In summary, the native gel confirmed that the Svp is catalyzing apo form Cpr36 to the holo form, which carries a different confirmation and is consistant with previous data.

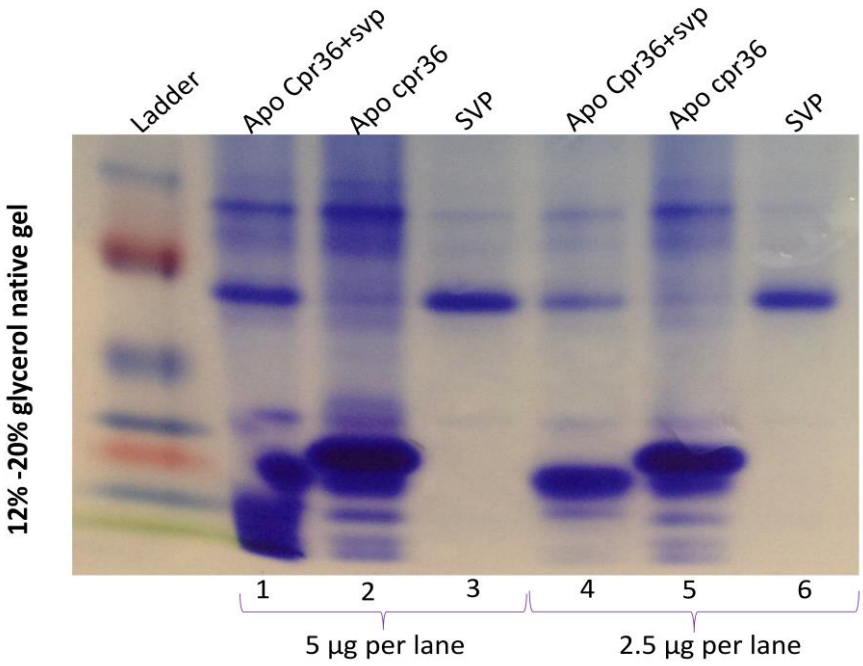

Figure 3.12 Nondenaturing polyacrylamide gel electrophoresis of Cpr36. 20% polyacrylamide gels were run at pH 8.5 and stained as previously described. Lanes: 1&4, apo form of Cpr36; Lane 2&5, apo form of Cpr36 treated with Svp; Lane 3&6, Svp enzyme.

Cpr36 was identified and characterized as an aryl carrier protein. HPLC, LC/MS, and ESI-MS data revealed that the apo form of Cpr36 was activated to form the holo form by the addition of 4′-PPT prosthetic groups to the conserved serine residue. The unexpected random mutation of Cpr36 from Met 63 to Val 63 had not change the aryl carrier protein function at all and the mutated apo-Cpr36 can still be transformed to the holo form. These results demonstrated that the activation of the carrier protein is highly site dependent and only the conserved serine residue can be adducted with 4′-PPT prosthetic groups.

Gene cloning and overexpression of *cpr37* from *E. coli* BL21 (DE3) has been found to be insoluble (data are not shown). The genes encoding *cpr37* were amplified from genomic DNA by PCR by using Herculase II Phusion DNA polymerase and digested with NdeI-BamHI and ligated to the identical sites of pXY200 to yield pXY200-cpr37 with expected size (Figure 3.13 A&B). The authenticity of *cpr37* was confirmed by the colony PCR and DNA sequencing which confirmed 5# colony harbored wild type gene of *cpr37* (Figure 3.13A).

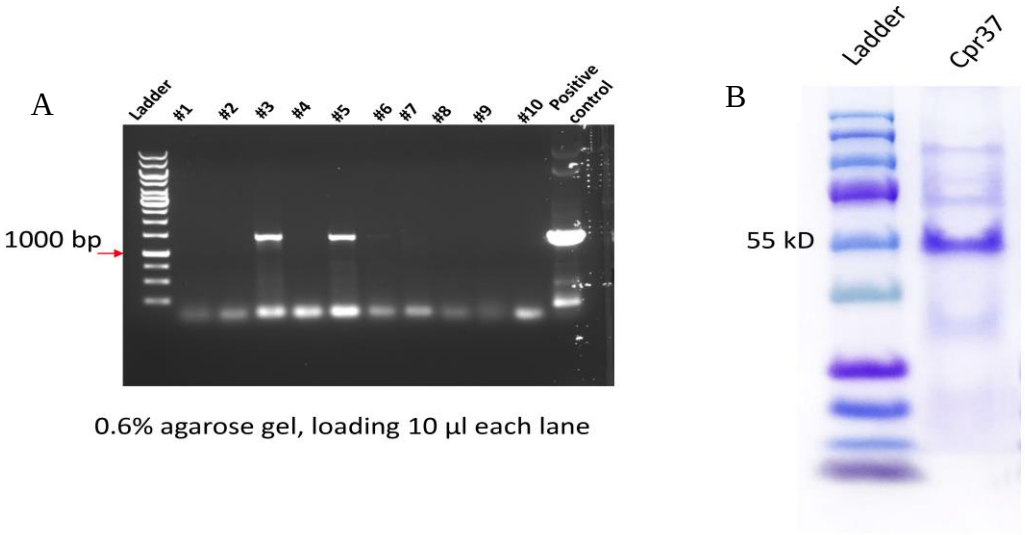

Figure 3.13 Overexpression and purification of Cpr37. (A) Colony PCR of pXY200-cpr37, only #3 and #5 colony shows the expected band. (B) SDS analysis of recombinant pXY200-cpr37 (expected MW of 55 kD).

The function of the adenylation domain of Cpr37 was identified using the carboxylic acid–dependent ATP-[^{32}]PPi exchange assay as described above (Figure 3.14). There was distinguishable difference between the experimental group (with Cpr37) and negative control (without Cpr37), providing more evidence that Cpr37 functions as the adenylation domain protein. The protein was purified from the *Streptomyces lividans TK-24* strain and inevitably included some protein impurities which were still visible from the SDS PAGE gel (Figure 3.13). To exclude the possibility of protein contamination, we conducted the carboxylic acid–dependent ATP-[^{32}]PPi exchange assay using PABA and different types of protein which were purified from the *Streptomyces lividans TK-24* cells only harboring intact pXY200 plasmid and harboring pXY200-Cpr37. Figure 3.14 shows distinguished difference between radioactivity between the two groups using different types of proteins, which exclude the possibility of protein contamination (Figure 3.15).

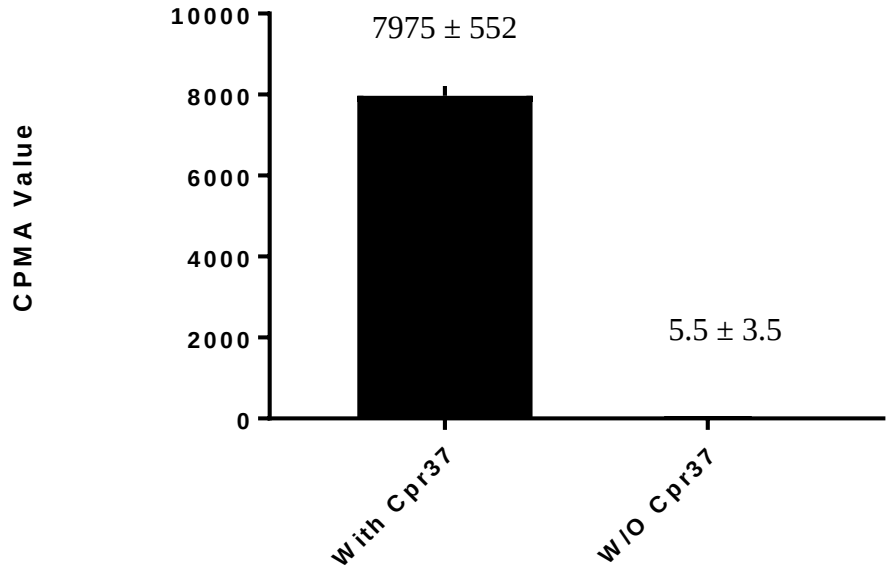

Figure 3.14 ATP-[32P]PPi exchange assay with PABA and Cpr37.

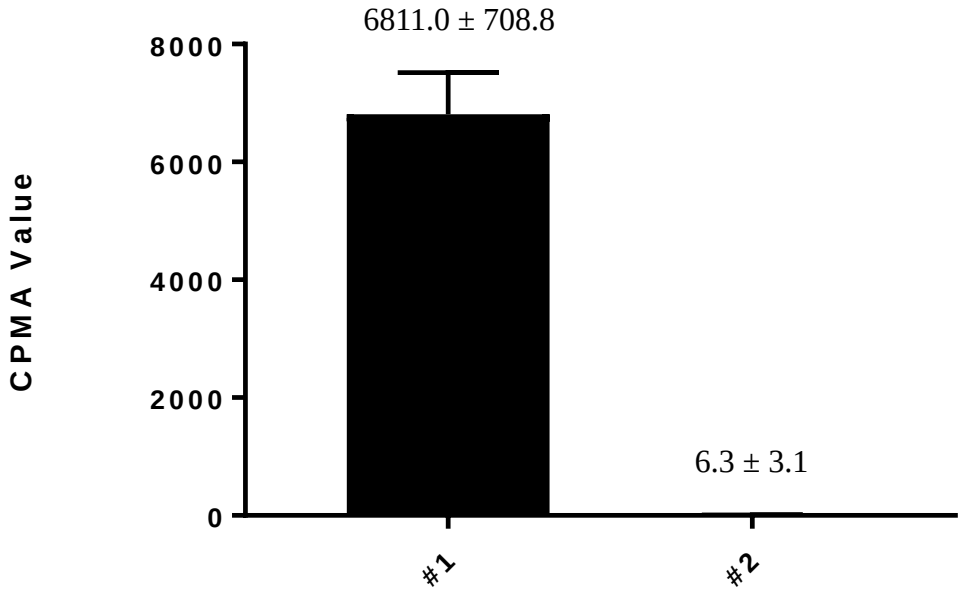

Figure 3.15 ATP-[^{32}P]PPi exchange assay with PABA and different proteins.
#1, proteins purified from the *Streptomyces* lividans TK-24 strain harboring pXY200-Cpr37. #2, proteins purified from the *Streptomyces* lividans TK-24 strain only harboring pXY200.

To better exploit the preference of different substrates, 20 proteinogenic amino acids, benzoic acid and PABA were tested by using the carboxylic acid–dependent ATP-[^{32}]PPi exchange assay. Although several amino acids were activated by Cpr37, including benzoic acid, phenylalanine, tyrosine, tryptophan and so on, the preferred substrate was PABA since all remaining amino acids had <60% relative activity (Figure 3.16). Notably, the amino acids which still show weak relative activity all share similar aromatic structure. Thus, these data have provided more evidence that Cpr37 acts as an adenylation domain protein. By introducing the hypothesized downstream reagents (Cpr36, PABA, SVP, [^{32}P]PPi, and CoA) into the reaction. The CPMA value decreases dramatically by comparing the data of Figure 3.14 (Figure 3.17), suggesting the downstream reaction happened and these downstream reagents pushed the chemical reaction forward, not turning back. When the enzyme Cpr37 was omitted, the CPMA

value also decreased dramatically (Figure 3.17).

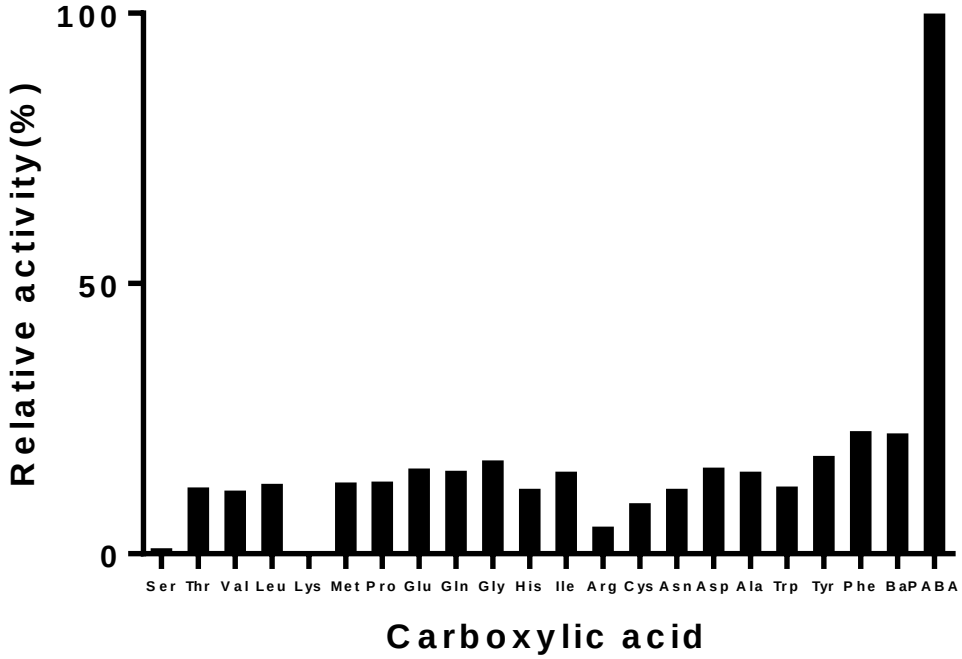

Figure 3.16 Amino acid-dependent ATP-[^{32}P]PPi exchange assay with Cpr37 and representative substrates (Ba is short for benzoic acid).

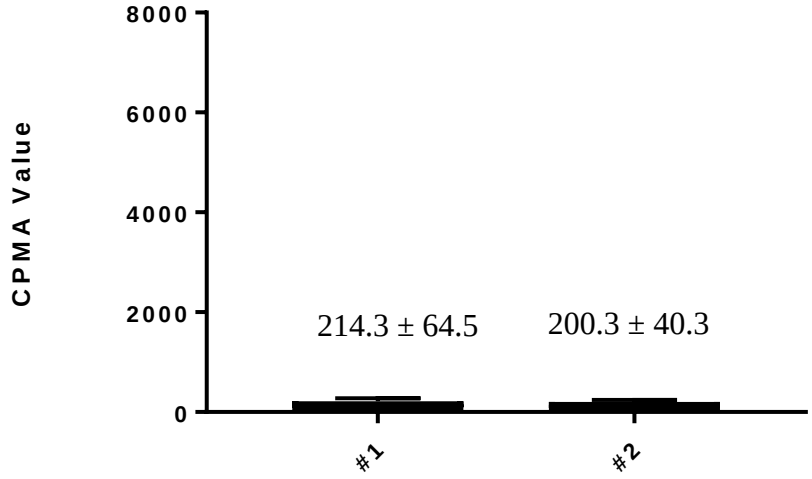

Figure 3.17 ATP-[^{32}P]PPi exchange assay with different groups of reagents. #1, reagents group of Cpr36, PABA, Svp, [^{32}P]PPi, and CoA. #2, reagents group of Cpr36, Cpr37, PABA, Svp, [^{32}P]PPi, and CoA.

In a traditional NRPS pathway, A domains release both AMP and pyrophosphate (PPi) throughout the catalytic progression. However, AMP is released only after the PCP-loading step, which makes PPi a more attractive target for monitoring the A domain activity and selectivity. Unluckily, detailed analyses of adenylation reactions by detecting PPi in a specific manner is not forthright. So we introduced the pyrophosphatase which can cut free inorganic pyrophosphate in aqueous condition, through which the kinetic study of Cpr37 can be achieved. After identifying the function of Cpr37, single-substrate kinetic experiments were performed using PABA as a substrate. Using nearly saturated ATP and variable PABA, the standard Michaelis-Menten kinetics were studied, yielding a K_m= 121 ±12 μM and k_{cat}= 2.32 ±0.2min (Figure 3.18). When some of the key reagents of the reaction were omitted, the obvious decrease of free inorganic phosphate concentration was observed (Table 3.2).

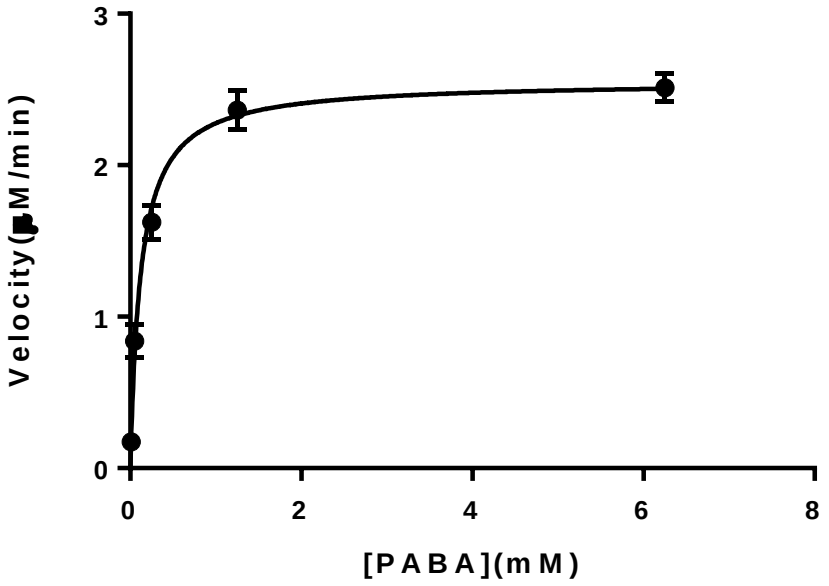

Figure 3.18 Plots for single-substrate of PABA kinetic analysis of Cpr37.

Table 3.2 Phosphate concentrations under different reaction conditions.

	ATP	PABA	Cpr37	phosphate (μM)
Reaction#				
1	+	+	+	18.69±0.26
2	−	+	+	$(1.46\pm1.19)\times10^{-1}$
3	+	−	+	5.19±0.77
4	+	+	−	$(6.61\pm1.45)\times10^{-1}$
5	+	−	−	$(8.07\pm1.29)\times10^{-1}$
6	−	+	−	$(5.14\pm3.57)\times10^{-1}$
7	−	−	+	$(5.28\pm3.64)\times10^{-1}$

Previous data have shown that the Cpr37 enzyme catalyzes the hydrolyzation of ATP to form PPi, which is illustrated by [^{32}P]PPi exchange assay. To provide more evidence about the incorporation of PABA into the carrier protein, [^{3}H] labeled PABA was applied to trace the amalgamation of PABA. Figure 3.19 revealed that PABA had been incorporated into Cpr36 and there was distinguished difference of [^{3}H] radioactivity for the groups treated with and without Cpr37. Following tracing experiment at different time points (Figure 3.20) also showed a steady increase of amalgamation of PABA. All these data supported that PABA is incorporated into Cpr36.

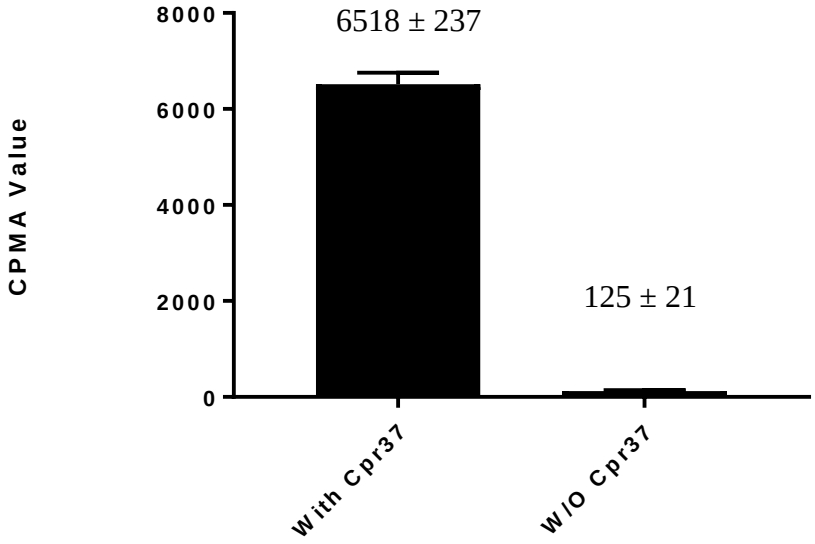

Figure 3.19 Plots for the incorporation of PABA by catalyzing Cpr37.

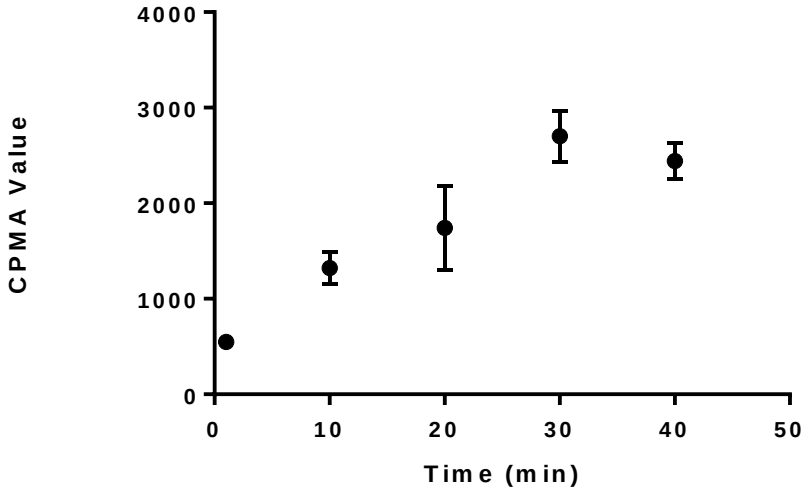

Figure 3.20 Plots for the incorporation of PABA at different time points.

To investigate the formation of thioester-linked PABA, HPLC was applied for the analysis. In the reactions including Cpr36, Svp enzyme, the retention time of peak assigned to holo form of Cpr36 protein showed ~0.5 minute later compared to the peak of original intact form of Cpr36. However, no obvious retention time difference was

observed between the thioester linked PABA Cpr36 and β-ketothioester linked PABA Cpr36. Further assays of HPLC/ESI MS–analyzed enzyme reactions provided more insights into the reaction pathway. Analysis of an enzyme coupled reaction by incubating Cpr36, Cpr37 and Svp with PABA yielded an $(M+Na)^+$ ion at molecular mass of 14904 Da, consistent with the molecular mass of PABA thioester linked Cpr36 (predicted mass , 14902 Da). Sodium was possibly introduced from ESI solvents or even solvents used during the protein purification process. The small molecule cleavage assay in chapter 4 provided more evidence that PABA was covalent attached to the 4′-PPT arm of holo-Cpr36 by detecting mild chemical release and analysis of intermediates bound to carrier proteins after treating with extra cysteamine (Figure 4.15 and 4.16).

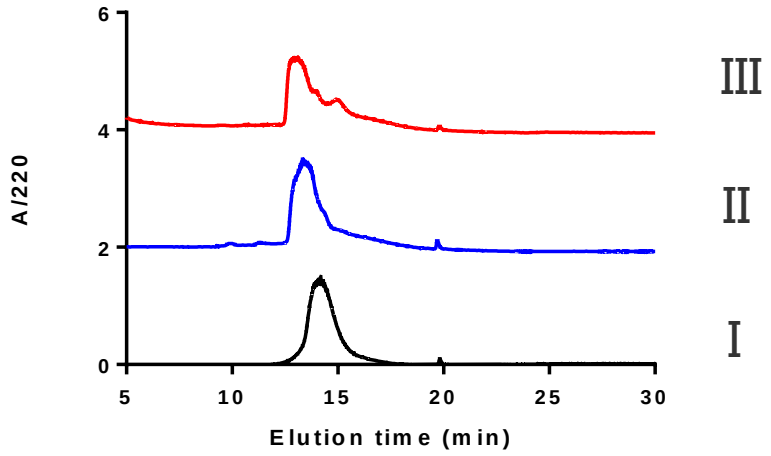

Figure 3.21 HPLC analysis of different Cpr36 derivatives. I, apo wild type Cpr36 as purified from *E. coli*; II, holo form of wild type Cpr36 resulted from Svp catalyzing by adding 4′-PPT arm from CoA into apo wild type Cpr36; III, apo wild type Cpr36, Svp, PABA and Cpr37.

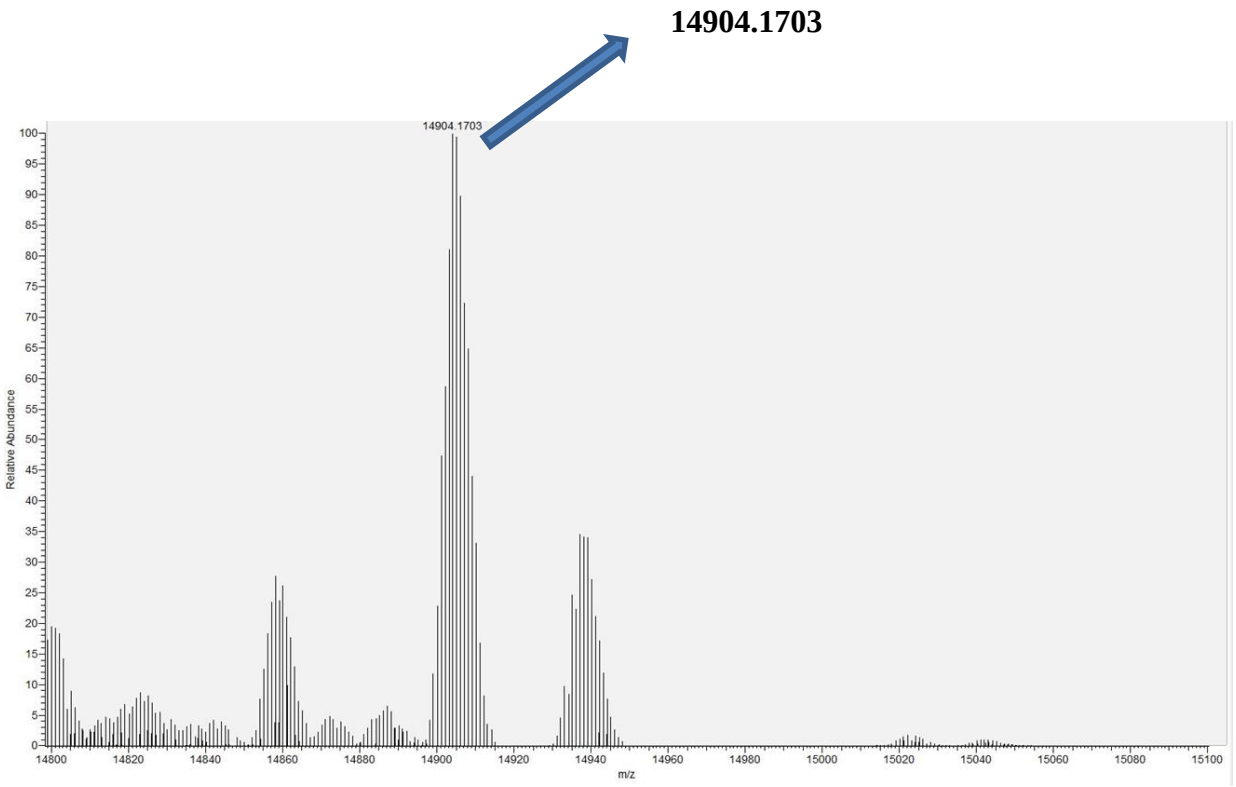

Figure 3.22 Molecular mass of PABA loaded carrier protein Cpr36.

3.4 Conclusion

In summary, functional characterization of Cpr36 and Cpr37 indicated that they catalyze identical reactions as AcmD and AcmA, suggesting a shared biosynthetic paradigm in antibiotics of actinomycin [201].

We have functionally and kinetically characterized Cpr36 as an aryl carrier protein. Similar with other aryl carrier proteins, it can be activated when Svp adds a 4′-PPT arm into its highly conserved serine residue. Although there was an unexpected mutation in the first batch of Cpr36 we purified, such mutation have no effect on the activation of the aryl carrier protein. The remaining steps are initiated by Cpr37-catalyzed activation of PABA as the acyl-adenylate with loading to the free-standing carrier protein Cpr36 to form the thioester-linked PABA.

Chapter 4: Elucidating the function and mechanism of Cpr33, Cpr34 and Cpr35

4.1 Background

With a stunning range of functional and structural diversity and medicinally important activities, the polyketide natural products are regarded as a class of compounds with remarkable properties, including antibiotic, anticancer, antifungal, antiparasitic and immunosuppressive [202]. Since biomimetic syntheses could provide indirect insights into the reactions occurring in vivo, there was tons of reasons to investigate the biosynthetic enzymes directly. The pathways of polyketide biosynthesis have strong homologies with the ones of fatty acid biosynthesis which are biosynthesized from acyl CoA precursors by polyketide synthases (PKSs) [203]. Both pathways apply the nature of the chemistry in chain extension from a shared pool of precursors and use the similar character of the enzymes for chain assembly. Truthfully, research in the fatty acid field has almost always guide researchers equivalently as in the field of polyketide biosynthesis.

The chemical steps of chain extension of fatty acids are shown in Figure 4.1. With a repeated head-to-tail linkage pattern, fatty acids are assembled from C2 units until reach required the length of a chain.

A starter acetyl unit is condensed with a malonyl unit that undergoes concerted decarboxylation and the resulting β-keto ester is successively reduced to a hydroxy, dehydrated and finally reduced again to give a longer saturated chain. A set of enzymes are used for chain extension. The starter acyl is initially attached to a cysteine thiol of a keto synthase (KS) and the chain extender malonate unit is bound to a thiol residue of a protein designated as an acyl carrier protein (ACP). In following steps, the keto ester is reduced by a keto reductase (KR), dehydrated by a dehydratase (DH) and finally reduced again by an enoyl reductase (ER). With the completion of one round of chain extension, the cycle is then repeated until leading to a chain of the required length. At the final stage, the chain is passed to a thioesterase (TE) enzyme from which it is

released as a free acid or an acyl ester. Besides the set of components (KS, ACP, KR, DH, ER and, TE), there is also a seventh partner, a malonyl-acetyl transferase (MAT). The role of MAT is to transfer the building blocks acetate and malonate from the respective coenzyme A pools onto the appropriate domains.

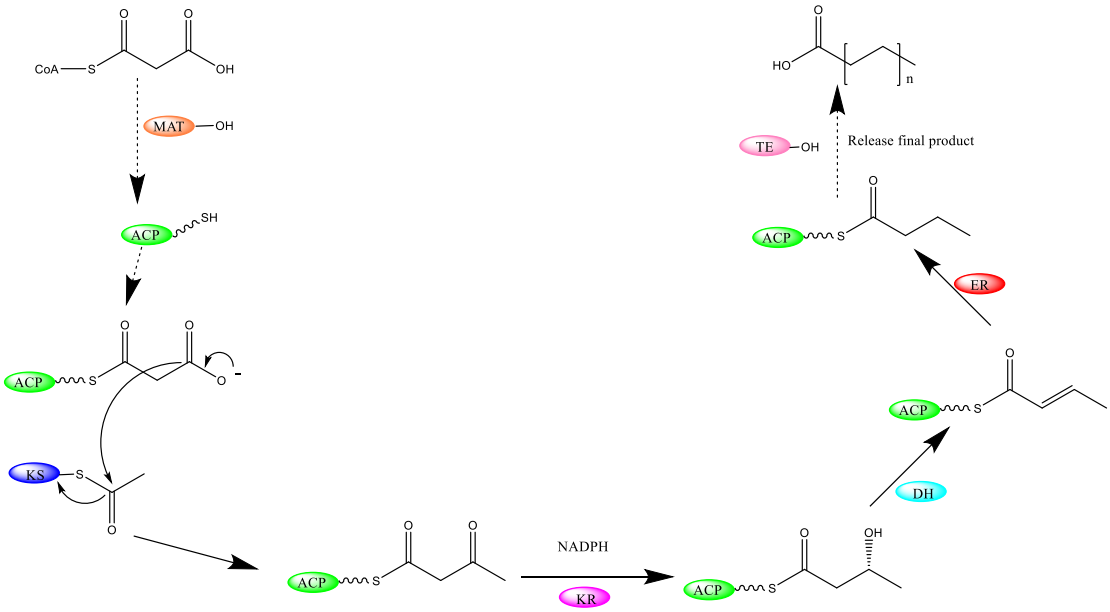

Figure 4.1 The fatty acid biosynthetic cycle. ACP: acyl carrier protein; MAT: malonyl-acetyl transferase; KS: ketosynthase; KR: ketoreductase; DH: dehydratase; ER: enoyl reductase.

The structural organization of the FASs varies from different types of organism. For example, in bacteria, a set of discrete proteins that can be isolated separately designated as the fatty acid synthase. However, in mammals, all of the enzyme activities are assembled to form a large multifunctional protein. Different intermediate stages of organization are found in other entities. The multifunctional variety are belonged to type I polyketide synthases. Contrariwise, the fully dissociable proteins are designated as type II polyketide synthases. Both systems are organized into a highly structured multifunctional arrangement. In the past several decades, type I polyketide synthases have attracted countless interest in the areas of understanding their structure and

function [204, 205]. However, only type II polyketide synthase (PKS) systems was proposed to recruit fatty acid biosynthetic machinery during the polyamide assembly of A-102395. As a consequence, we will only present a comprehensive introduction of all of the advancements in the field.

Simpson et al. have determined the structure and function of the ACP components of a series of minimal PKSs [206-208]. The full solution structure of the actinorhodin ACP shows the remarkable similarity to the fatty acid ACP derived from *E. coli* with one exception that the putative chain binding pocket in the polyketide ACP is less lipophilic [209, 210]. Bristol group discovered type II polyketide ACPs can accept a malonate and other acyl groups from their respective CoA thioesters apparently without external assistance [206, 211]. Recently, Kevin Reynolds further proved that a type II ACP is not just an ACP, but is also capable of a powerful AT activity [212]. In the iterative type II PKSs, there is only one set of heterodimeric KS (KS_α-KS_β) that have to operate a specific number of times in building a polyketide chain in correct length [213]. Another important discovery about type II PKS systems was that Leadlay found KS_β showed the remarkable sequence homology to its partner, KS_α. However, he also noticed that cysteine was missing in the active site of KS_β and the missing residue was not random but was always glutamine. Therefore this replacement might cause KS_β to act as a decarboxylase toward malonate [214].

C. Bisang et al. found that a minimal mutant PKS in which the KS_α active site cysteine was substituted for alanine disabled polyketide biosynthesis. Analysis of the resulting KS_α/KS_β complex showed that the malonyl-ACP was certainly readily decarboxylated to acetyl-ACP (Figure 4.2(a)). Similarly, no decarboxylase activity was observed in another mutant in which both the KS_α/KS_β active site residues were changed to alanine. Together, these results unambiguously demonstrated the crucial role of active site glutamine of KS_β for aromatic polyketide assembly [214]. Under the condition of replacing cysteine with a glutamine at the active site of a condensing KS, Stuart Smith et al. demonstrated that the KS could be transformed into a decarboxylase [202].

The biosynthesis of aromatic polyketides follows the sequence of events as following. KS_{β}, functions as a chain initiation factor, catalyzing the decarboxylation of malonyl-ACP to form acetyl-ACP (Figure 4.2(a)). Once decarboxylation of malonyl-ACP, the acetyl group is transferred to the KS_{α} active site and the ACP is malonylated again and proceeds to form acetoacetyl-ACP (Figure 4.2(b)). This ketoester is then transferred from the acetoacetyl-ACP to KS_{α} again, which initiate another cycle of chain extension [202].

The mechanism responsible for governing chain length hasn't been discovered yet due to some challenges. Firstly, there is some difficulty of the chemical analysis during the investigating this question. PKS products with very similar properties are produced often by mix-and-match experiments. Additionally, there is a high risk that polyketide metabolites isolated from a complex mixture of materials in fermentation media will not be detected [202]. The other factor that determines chain length is the release of the fully extended molecule. Peter Shoolingin-Jordan proposed one possible mode of release for type II PKS systems through a ketene mechanism [215]. Some key agent in the PKS complex can determine the timing of release and formation of an α-pyrone would be a very favorable chemical mechanism for chain release [202].

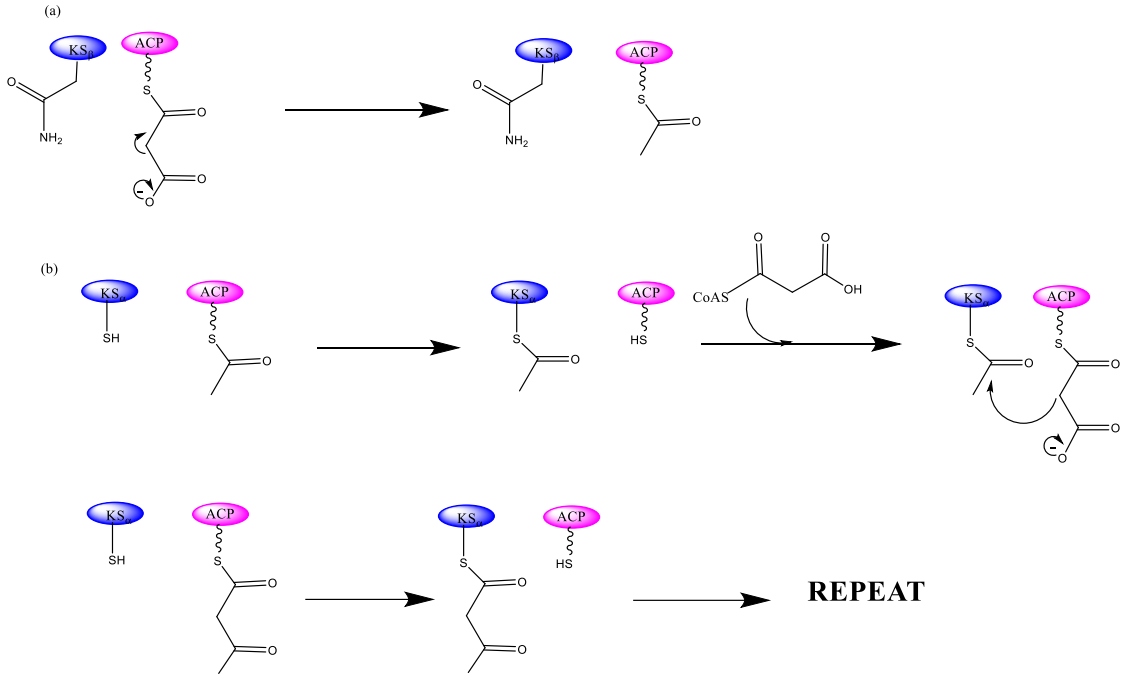

Figure 4.2 Sequence of events in the biosynthesis of aromatic polyketides. (a) The KS_β domain initiate chain extension through decarboxylation of malonyl-ACP to acetyl-ACP. (b) The acetyl group is then passed to the active site cysteine of KS_α. KS-catalysed condensation takes place between the malonate loaded ACP and the activated KS_α, resulting in acetoacetyl-ACP. The ketoester can be transferred back to the KS_α, and another cycle of chain extension begin.

PKS structure study is one area that attracts enormous interest. Although X-ray crystallography and NMR have already yielded high-resolution structures of individual PKS domains, the nature of the quaternary association is still little known. Lack of such kind of knowledge has contributed to the failure of mix-and-match experiments to yield predictable products [202].

Bioinformatic sequence analysis illustrates that Cpr33 shows closest similarity to 3-oxoacyl-ACP reductase (FabG) from bacterial type II fatty acid synthases, which catalyze hydride addition to the si face to generate the (3R)-3-hydroxyacyl-[acyl-carrier-protein] [106]. Specifically, 3-oxoacyl-ACP reductase catalyzes the chemical reaction as follows:

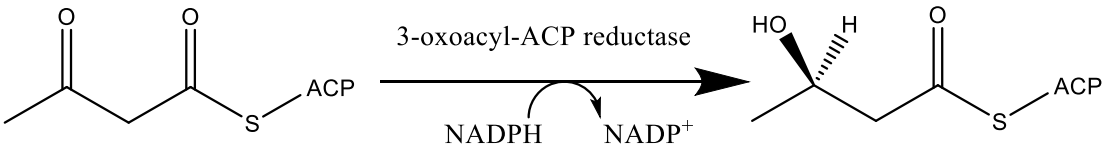

Figure 4.3 Reaction catalyzed by 3-oxoacyl-ACP reductase.

Therefore, the substrates of this enzyme are 3-oxoacyl-[acyl-carrier-protein], NADPH, and H^+, whereas products are (3R)-3-hydroxyacyl-[acyl-carrier-protein] and $NADP^+$. FabG is stereochemical workhorses which set the specific stereocenters within polyketide scaffolds.

KRs exert quite exquisitely stereocontrol since they can be both stereoselective and stereospecific. KRs can be stereoselective in generating either an L-β-hydroxyl group polyketide intermediates (A-type KRs) or a D-β-hydroxyl group polyketide intermediates (B-type KRs). In the biosynthetic pathways of fatty acids and polyketides, ketoreductions are performed by shortchain dehydrogenase/reductase (SDR) enzymes on β-ketoacyl substrates to generate D-β-hydroxyacyl products. PKS KRs also belong to SDR enzymes, and B-type KRs execute equivalent stereoselective reductions. All these enzymes share following features: (1) similar nicotinamide coenzyme binding sites and catalytic residues. (2) three glutamine residues before the catalytic tyrosine and an aspartate on a loop adjacent to the active site which is referred to as "Loop DE". Adrian proposed a mimic model of the natural substrate based on the acetoacetyl-CoA reductase complex. They found that a crystal structure of the B-type KR MlnKR6 bounded to $NADP^+$ shows a comparable interaction between a well-ordered water molecule and the Leu-Asp-Asp motif. A hydrogen bond is formed between the β-keto group and the catalytic tyrosine. In such scenario, reduction by NADPH will stereospecifically yield a β-hydroxyl group with a D-orientation [216].

In the B-type KRs such as FabG, the aspartate orients the glutamine. Stereochemistry may be controlled through the hydrogen bond which was donated by the NH_2 group of a conserved glutamine to the substrate thioester carbonyl. In enzymes that catalyze an

‘‘R’’ form of reduction, NH_2 group of glutamine is oriented by a conserved aspartate may so that a polyketide enters from the right side of the groove. However, in enzymes catalyzing the ‘‘S’’ reaction, the glutamine NH_2 is oriented by a tryptophan on the left side of the groove so that the polyketide enters from the left side [217].

In *E. coli*, an open reading frame encoding a protein with strong similarities to particularly plant ketoacyl-ACP reductases with over 50% identical residues is located within the *fab* gene cluster designated as *fabG* between the *fabD* and *acpP* genes [218]. There is only a single NADPH-specific β-ketoacyl-ACP reductase in *E. coli* which functions with all chain lengths [219]. Hence, fatty acid synthesis following the initial condensation should be blocked in *fabG* mutants. In conjunction with the cofactor NADPH, FabG catalyzes the reduction of the carbonyl group at the C3 position to a hydroxyl group by the addition of two hydrogen atoms, one to the carbonyl oxygen and the second to the carbonyl carbon.

There is only a single isozyme of FabG known in nature [220]. These considerations suggest that FabG may be an ideal target for the development of broad-spectrum antibacterials. It is found that there is only a single report of FabG inhibitors, however, these compounds are not suitable scaffolds for drug discovery [221]. It turned out that the active site tunnels in FabG only form in the company of the acyl-ACP substrate bound to the enzyme [222].

So far, there are three conformations of FabG from *E. coli* determined, which include the protein structure without cofactor, the EcFabG·$NADP^+$ binary complex, and an EcFabG[Y151F]·$NADP^+$ mutant structure [222, 223].

4.2 Materials and Methods

4.2.1 Instrumentation, Chemicals and Bacterial Strains

NADPH, isopropyl-β-Dthiogalactopyranoside (IPTG), cysteamine, nucleoside bases, were purchased from Sigma (St. Louis, MO). Buffers, salts, and media components were purchased from Fisher Scientific. Synthetic oligonucleotides were purchased from

Integrated DNA Technologies (Coralville, IA) and are listed in Table 4.1. Genomic DNA from *Streptomyces* sp. A-102395 was extracted using UltraClean Microbial DNA Isolation Kit (MoBio laboratories, Inc) following the manufacturer's protocol. DNA quality and concentration were confirmed using Nanodrop 2000c spectrophotometer (Thermo Scientific) and gel electrophoresis.

All NMR data was recorded at 500 MHz or 400 MHz for ^{1}H and 100 MHz for ^{13}C with Varian Inova NMR spectrometers (Agilent, Santa Clara, CA). LC-MS was conducted with an Agilent 6120 Quadrupole MSD mass spectrometer (Agilent Technologies, Santa Clara, CA) equipped with an Agilent 1200 Series Quaternary LC system and an Eclipse XDB-C18 column (150 × 4.6 mm, 5 μm). HR-ESI-MS spectra were recorded on an Agilent 6230B Time of Flight (TOF) LC/MS (Agilent Technologies, Santa Clara, CA, USA). Analytic HPLC was performed with one of two systems: a Waters Alliance 2695 separation module (Milford, MA) equipped with Apollo C4 column (250 x 4.6 mm, 5 μm) purchased from Grace (Deerfield, IL). All solvents and medium used were purchased from the same manufacturer as above mentioned.

4.2.2 Cloning, Overexpression and Purification of Proteins of Cpr34, Cpr35 and Cpr33

The genes encoding Cpr34, Cpr35 and Cpr33 were amplified from genomic DNA extracted from *Amycolatopsis sp.* SANK 60206 by PCR using Herculase II Fusion DNA Polymerases from Agilent Genomic with supplied buffer , 200 μM dNTPs, 5% DMSO, 10 ng template, 10 U DNA polymerase, and 200 nM each of the primer pairs in table 4.1.

Table 4.1 Primers for amplification of *cpr34*, *cpr35*and *cpr33*.

Name	Sequence
cpr34_forward	5′- GAT ACGCATATG GTGCATGAGGTCGTTGTAACAGG-3′
cpr34_reverse	5′- TAGAGTGGATCCTCAACGGGAAGGTGGAGC-3′
cpr35_forward	5′-GATACGCATATGATGCCGTCAGCGCTCAT-3′
cpr35_reverse	5′- TAG AGT GGATCCTCACAGGGAAAGCTCGTAGATG-3′
cpr33_forward	5′-GGTATTGAGGGTCGCATGATCACTGTTGAGGATGT CAAG -3′
cpr33_reverse	5′-AGAGGAGAGTTAGAGCCGTGCCGCGTTCAGTATT GGT-3′

The gel-purified PCR product of *cpr33* was inserted into pET-30 Xa/LIC using ligation-independent cloning as described in the provided protocol to yield pET30-*cpr33*. The gel-purified PCR product of *cpr34* and *cpr35* were digested with NdeI-BamHI and ligated to the identical sites of pDB.His.MBP to yield pDB.His.MBP-cpr34, pDB.His.MBP-mur35. The identity of the cloned genes was confirmed by DNA sequencing. Plasmids pDB.His.MBP-*cpr34*, pDB.His.MBP-*cpr35* and pET30-*cpr33* were introduced into *E. coli* BL21 (DE3) cells, and the transformed strains were grown in 5 mL LB culture supplemented with 50 μg/mL kanamycin overnight at 37 °C. The recombinant strains were grown in 500 mL of LB with 50 μg/ml kanamycin until the cell density reached an OD600~0.5, then 0.1 mM 1-thio-β-Dgalactopyranoside (IPTG) was added to induce expression. Cells were collected by centrifuge at 4°C for 10min then resuspended in lysis buffer containing 100 mM Tris-HCl (pH=8.0), 300 mM NaCl and 10 mM imidazole (pH 8.0). Lysis of cells was performed using a Qsonica sonicator (Qsonica LLC, Newtown, CT) for sonication for a total of 6 min at 40% amplitude with 30 s pulses separated by 30 s rest periods. The culture was kept on ice. Immediately after lysis. Cells debris were separated from lysate by centrifuge at 10,000 rpm for

60min. The supernatant containing His6-tagged protein was loaded onto Ni-NTA agarose (Thermo Scientific, Rockford, IL) and washed with 50 mL of washing buffer containing 100 mM Tris-HCl (pH=8.0), 250 mM NaCl and 20 mM imidazole. The recombinant proteins were eluted with 10 mL of elution buffer containing 250 mM NaCl, 100 mM Tris-HCl (pH=8.0) and 200 mM imidazole, and then concentrated with an Amicon Ultra 10000 MWCO centrifugal filter (Millipore), followed by desalting of the proteins using desalting buffer containing 50mM Tris-HCl (pH=8.0) and 100mM NaCl. The proteins were stored in 40% glycerol at -20 °C until use. Size and purity was assessed as by 12% acrylamide SDS-PAGE and the concentration was determined by Nano Drop according to the predicted extinction coefficient of each protein using the ProtParam tool available from ExPASY.

4.2.3 Reaction of Transformation from ACP to Malonyl-S-ACP

Reactions consisted of 10 mM Tris-Hcl(pH 7.5), 5 mM $MgCl_2$, 250μM Cpr36, 150 μM Svp, and 0.5mM malonyl-CoA, with the water up to the volume of 100μL at 30 °C for one hour.

4.2.4 Enzyme Reaction Catalyzing by Cpr34 and Cpr35

To determine the reaction catalyzed by Cpr34 and Cpr35, protocol were developed as following: reaction pool 1(100 μL) containing 50Mm Tris-Hcl (pH=8.0), 5mM $MgCl_2$, 200μM CoA, 100 μM Cpr36, 20 μM Svp and water up to 100 μL was incubated at room temperature for 1 hour until the apo-form of Cpr36 was converted to holo form completely, which can be determined by HPLC as previously described. Adding another pool 2 of reaction mixture which containing 200μM PABA, 200 μM ATP and 20 μM Cpr37 into the reaction system. Reaction pool 3 containing 10 mM Tris-Hcl (pH 7.5), 5 mM $MgCl_2$, 100μM Cpr36, 20μM Cpr34, 20μM Cpr35, 20 μM Svp, and 200μM malonyl-CoA, with the water up to the volume of 100μL at 30 °C for one hour was added to the previous reaction pool for overnight. The reaction was directly analyzed by HPLC equipped with Apollo C4 column (250 x 4.6 mm, 5 μm) purchased from

Grace (Deerfield, IL) using the same gradient programs described previously in chapter 3.

4.2.5 Protein Mass Spectrometry Analysis of Reaction Catalyzing by Cpr34 and Cpr35

The individual protein reactions peaks were collected from HPLC and conduct high-resolution mass spectrum characterization as previously decribed.

4.2.6 Enzyme Reaction Catalyzing by Cpr33

To study the function of Cpr33, 10 mM NADPH, 10 μM Cpr33, 100 mM KH_2PO_4, 300 mM NaCl and 1 mM EDTA was added into the reaction pool #3 described in 4.2.4 which consisted of 10 mM Tris-Hcl (pH 7.5), 5 mM $MgCl_2$, 100μM Cpr36, 20μM Cpr34, 20μM Cpr35, 20 μM Svp, and 200μM malonyl-CoA. Water was added up to the volume of 200μL and incubated at 30 °C for overnight.

4.2.7 Cysteamine-Promoted Cleavage of Enzyme-Bound Polyketides

A method was used to cleave the small molecule from the enzyme bounded polyketide as previously described [199]. Rather than the mechanism of direct nucleophilic attack by the nitrogen of cysteamine for the derivatization, following transthioesterification and the expected intramolecular rearrangement, the resulting free thiols of the polyketide adducts are oxidized to the mixed disulfides 1-3a with treatment of excess free cysteamine (Figure 4.4). This strategy introduces a free amine and lead to a dramatic increase in MS analytical sensitivity.

A 1 M solution of cysteamine-HCl (Fluka/Sigma-Aldrich, St. Louis, MO) was prepared in nickel resin elution buffer at neutral pH. Freshly purified reaction solutions of reaction pools #2 and #3 (as described in 4.2.4) and reaction pool #4 which consisted of all the same components in reaction pool 3# and the following ingredients: 10 mM NADPH, 10 μM Cpr33 and 1 mM EDTA in nickel elution buffer were supplemented

with the buffered cysteamine solution to a final concentration of 0.2 M cysteamine. The whole reaction tubes were capped, and wrapped in aluminum foil. Cysteamine treatment proceeded for 20 h at 4 °C. Samples were then extracted twice with ethyl acetate and concentrated by rotary evaporator. The residue was immediately resuspended in an appropriate solvent, filtered, and analyzed by HPLC/LC-MS and high resolution mass spectrometer using the gradient programs provided as follows. A series of linear gradients from 0.1% Formic acid in 100% water (A) to 0.1% Formic acid in acetonitrile (B) in the following manner (beginning time and ending time with linear increase to % B): 0-16 min, 5%-100% B; 16-20 min, 100% B; 20-30 min, 100%-5% B. The flow rate was kept constant at 0.4 mL/min, and elution was monitored at 254 nm.

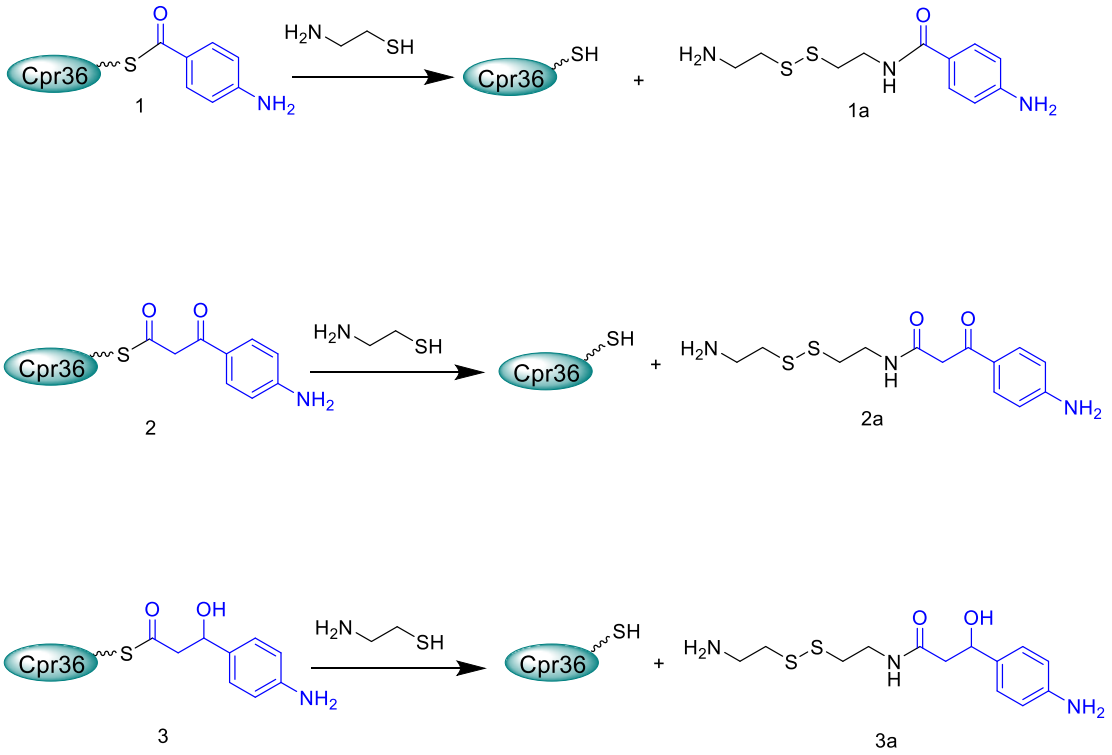

Figure 4.4 Reaction cascade leading to cystamine adducts1-3a.

4.2.8 Synthesis of the Substrate of Cpr33

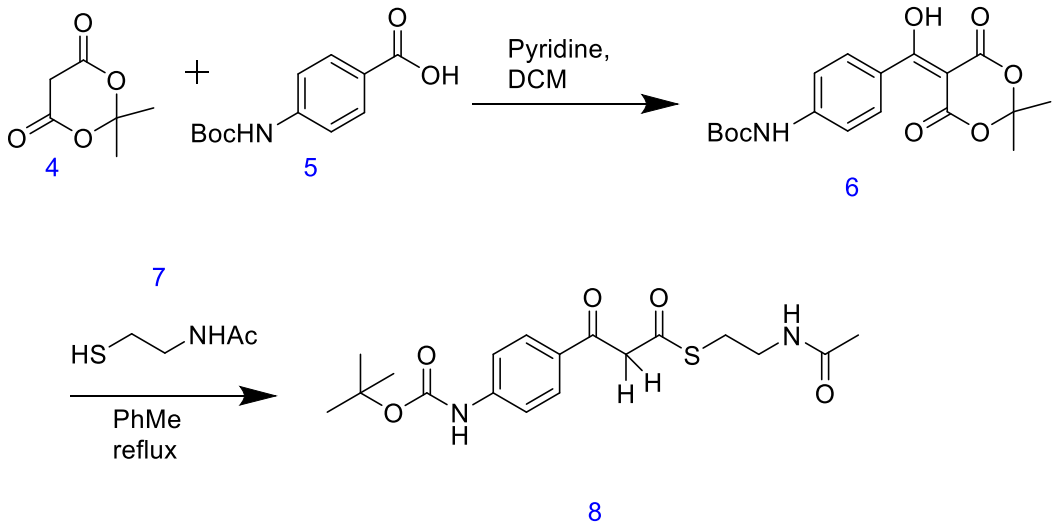

Figure 4.5 Chemistry synthesis scheme of tert-butyl N-(4-{3-[(2-acetamidoethyl) sulfanyl]-3-oxopropanoyl}phenyl)carbamate.

To better exploit the unique structural properties of the products of these capuramycin series enzymes, especially the stereochemistry properties of product of Cpr33, the mimic intermediate of beta ketoester linked Cpr36 was synthesized by following the synthesis scheme of Figure 4.5. Meldrum's acid (**4**) (1.0 g, 4.21 mmol) and Boc-Abz-OH (**5**) (0.67 g, 4.64 mmol) were dissolved in DCM (50 mL) in a reaction vessel under nitrogen atmosphere. DMAP (0.77 g, 6.32 mmol) was added to the mixture and cooled to 0 °C. A solution of DCC (1.03 g, 4.97 mmol) in DCM (30 mL) was added dropwise to the above mixture. The reaction mixture was allowed to warm to RT and stirred overnight. The slurry mixture was then poured into cold ethyl acetate and allowed to stirred for 30 mins in ice bath. The slurry was filtered and filtrated was washed with

cold 5% citric acid (2 times), cold water (2 times), and cold brine (2 times). The organic layer was dried over $MgSO_4$ and concentered under vacuum. The solid obtained was washed with petroleum ether (3 times) and dried under vacuum to yield 0.72 g (28.56 %, 1.98 mmol) of **6**. The compound **6** was taken to the next step without purification. The compound **6** (0.72 g, 1.98 mmol) was dissolved in dry toluene (7.5 mL) in a reaction vessel and N-acetylcysteamine (**7**) (0.24 g, 1.98 mmol) in dry toluene (7.5 mL) was added to it dropwise under nitrogen atmosphere. The mixture was stirred at 80 °C for 6.5 h and monitored through TLC for the consumption of reactants. Once reaction was completed, it was allowed to come to RT, and excess solvent was evaporated. It was poured onto a silica column, then eluted with ethyl acetate. The solvent was evaporated to yield 0.169 g (22.42%, 0.44 mmol) of **8**.

Purified **8** was analyzed by high resolution MS. It revealed a more convincing data with $(M + H)^+$ ion at *m/z* = 381.1488, which is consistent with the molecular weight of 8(Figure 4.7). 1D and 2D NMR, and relative peak intensities were assigned based on natural abundance ^{13}C and ^{1}H NMR spectra (Figure 4.8). Correlations obtained from COSY, NOESY, HMBC and HMQC experiments of **8** were used to establish the structure (Figure S.4.1-4.4).

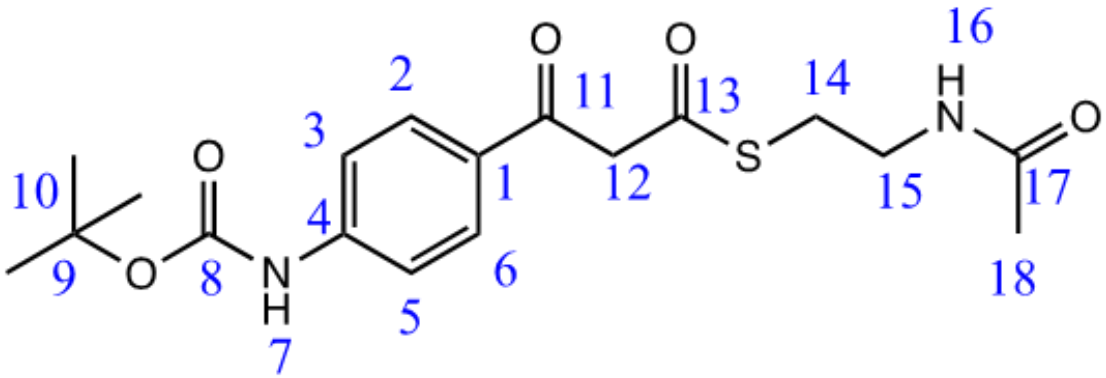

Figure 4.6 Number labeled structures of intermediate of **8** for NMR.

Table 4.2 NMR data of **8** in $CDCl_3$ (^1H: 400 MHz, ^{13}C: 100 MHz, J in Hz).

Positions	δ_C	δ_H
1	144.35	-
2	127.80	7.74 (d, J = 8.8 Hz, 1H)
3,5	23.2	0.48
6	130.42	7.90 (d, J = 9.0 Hz, 1H)
7	-	6.70 (s, 1H)
8	152.40	-
9	81.51	-
10	28.41	1.53 (s, 9H)
11	190.89	-
12	53.62	4.21 (s, 1H)
13	193.10	-
14	29.34	3.13 (dt, J = 13.7, 6.8 Hz, 2H)
15	39.38	3.68 – 3.31 (m, 2H)
16	-	5.95 (bs, 1H)
17	170.92	-
18	23.27	1.98(s,3H)

Further HRMS data of figure 4.7 also matches (*m/z* calculated for $C_{18}H_{24}O_5N_2S$ (M + H): 381.1479; found: 381.1488). All these combined data prove the authenticity of 8.

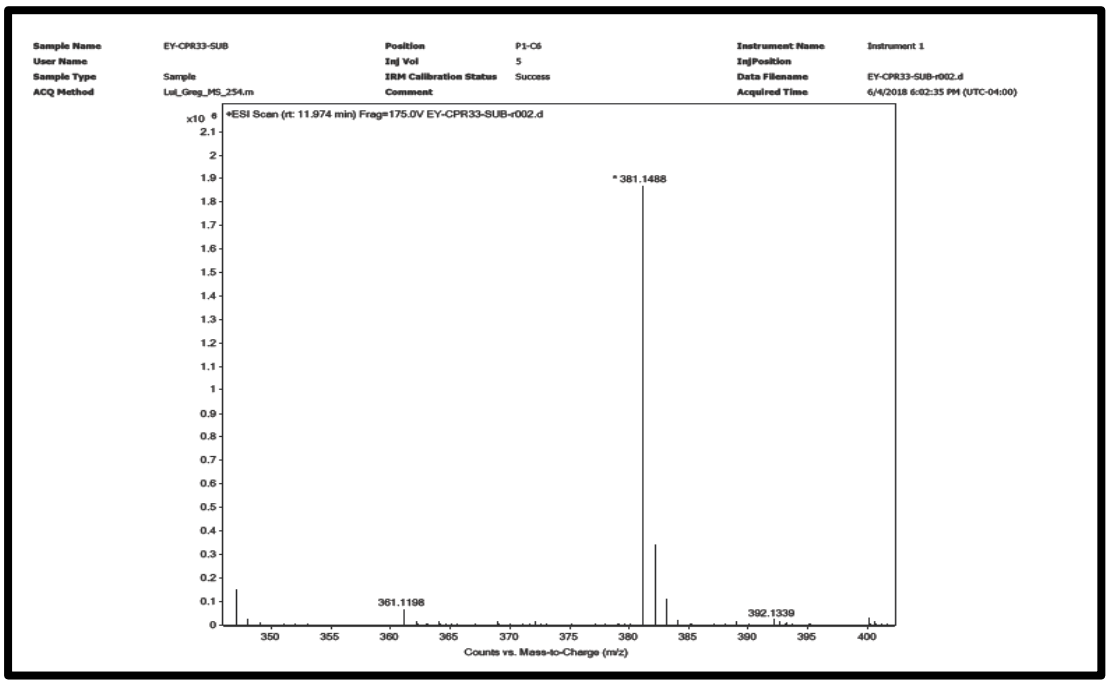

Figure 4.7 (+)-HR-ESI-MS (positive mode) of 8.

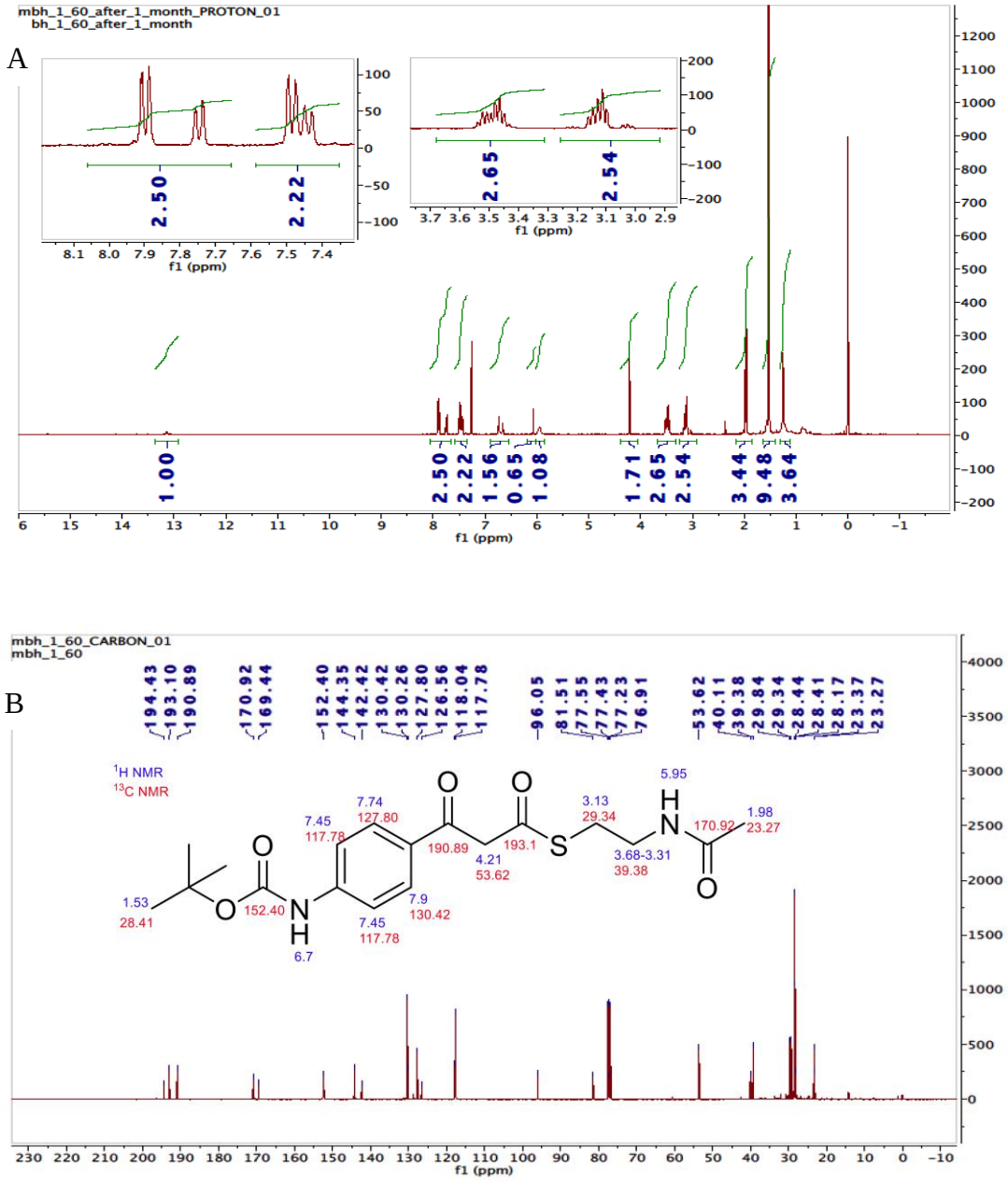

Figure 4.8 NMR spectrum of 8. (A) ^{1}H NMR spectrum ($CDCl_3$, 400 MHz) and (B) ^{13}C NMR spectrum ($CDCl_3$, 400 MHz).

4.3 Results and Discussion

Structural homology search using HHpred was carried out [106] and BLASTP analysis showed that Cpr34 and Cpr35 belong to ketoacyl synthase superfamily with Cpr34 resembled β-ketoacyl synthase and Cpr35 resembled α-ketoacyl synthase. Proteins in

ketoacyl synthase superfamily, catalyzing the decarboxylative condensation in the biosynthesis of polyketide and fatty acid, has diverse functions. Thus, we proposed that Cpr34 and Cpr35 work in concert to catalyze decarboxylative condensation between a malonyl-*S*-acyl carrier protein (ACP) and recipient thioester during biosynthesis of A-102395 (Figure 4.9). A comparable mechanism of condensation has previously been proposed during hygromycin B biosynthesis upon identification of its gene cluster [224].

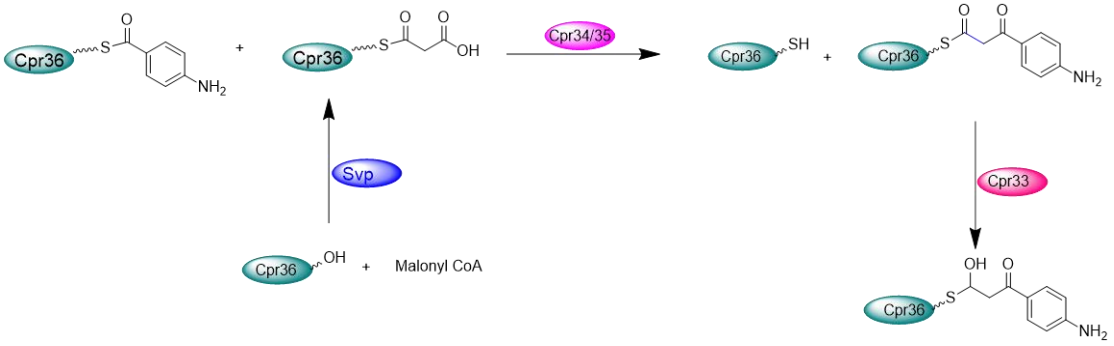

Figure 4.9 Porposed reaction catalyzed by Cpr34/35 and Cpr33.

The previous data have provided a lot of evidence that the Cpr36 can be activated by Svp with adding 4′-PPT arm into its highly conserved serine residue. To determine whether other CoA derivatives can also be loaded into apo form of Cpr36, C^{14} labled acetyl CoA and malonyl CoA were used to trace the incorporation of 4′-PPT arm. The results showed that there was distinguished radioactivity for the groups treated with the Svp. No radioactivity was observed when Svp was omitted in the same reaction (Figure 4.10). Further protein mass spectrum also provided more evidence that the 4′-PPT arm from the acetyl CoA and malonyl CoA was added to apo form of Cpr36 to finally form malonyl-Cpr36 since the molecule weight matched (Figure 4.1, expected mass for malonyl CoA-Cpr36, 14846 observed mass, 14847; expected mass for acetyl CoA-Cpr36, 14803, observed mass, 14802).

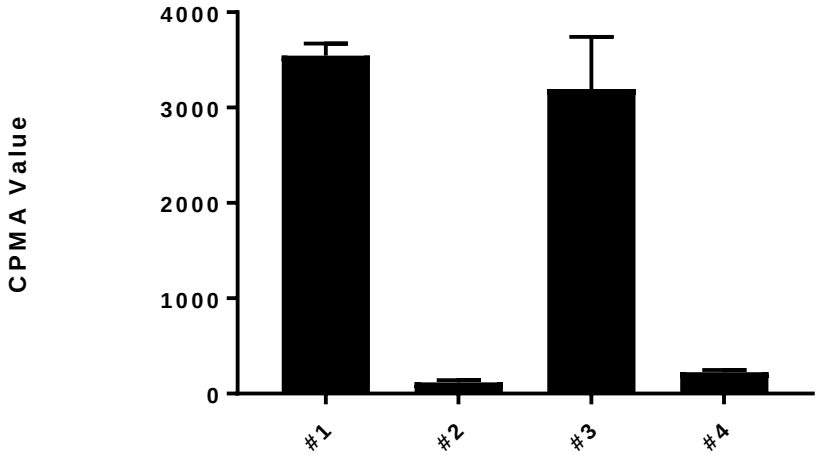

Figure 4.10 C^{14} labled malonyl CoA and acetyl CoA to trace the incorporation of 4′-PPT into Cpr36. #1, Cpr36, malonyl CoA and Cpr37. #2, Cpr36, malonyl CoA without Cpr37. #3, Cpr36, acetyl CoA and Cpr37. #4, Cpr36, acetyl CoA without Cpr37.

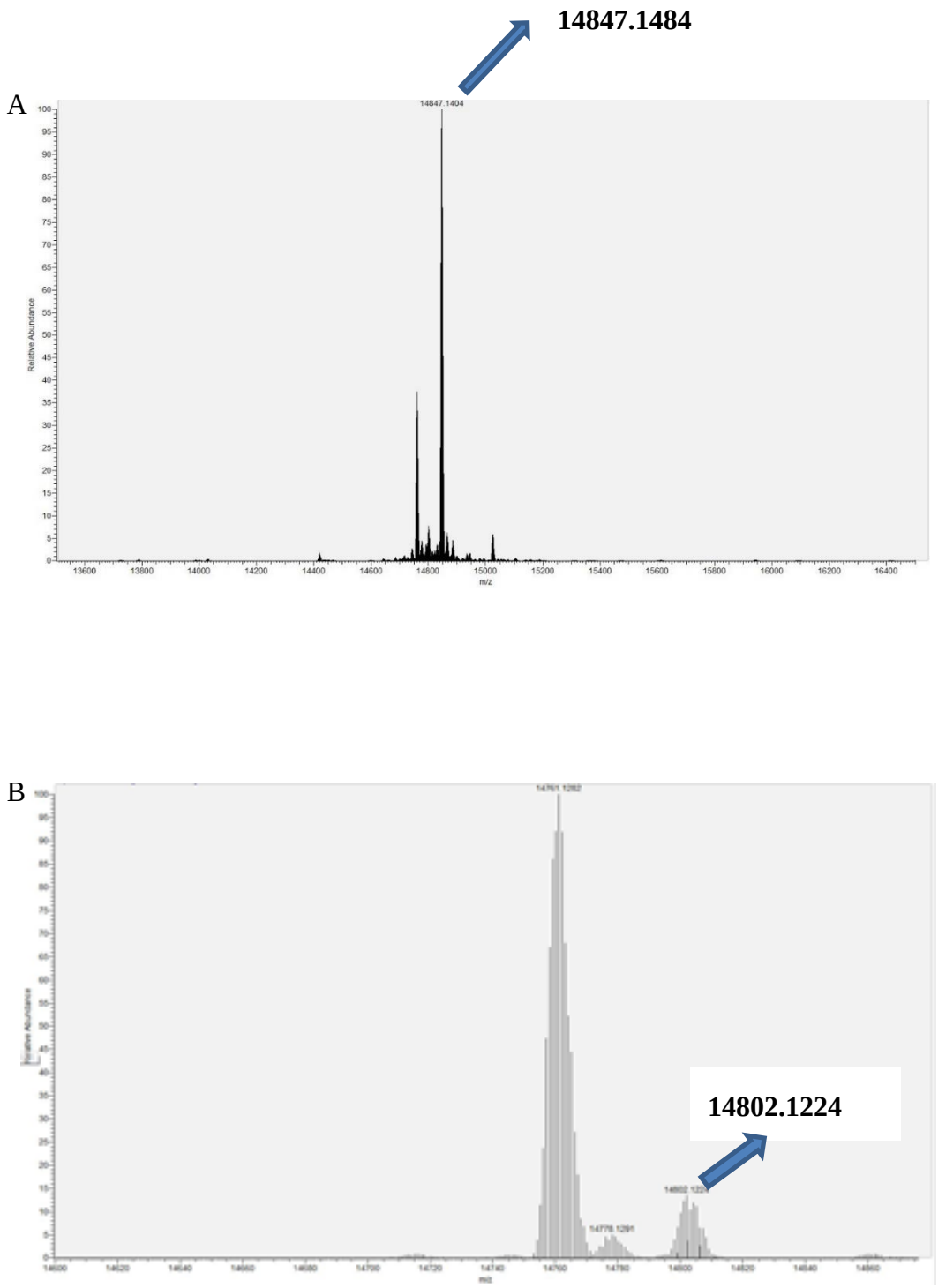

Figure 4.11 Molecular mass change of Cpr36 after loading different CoAs. HPLC/ ESI MS–analyzed enzyme reactions using an analytical HyPURITY™ C4 HPLC column (150 mm × 4.6mm, Fisher Scientific Co., Ltd.) reversed-phase at 25 °C, at a flow rate of 0.4 mL min–1. Asterisks denote degraded protein impurities. (A) Mass spectrum for the protein peak eluting at time t ≈14 min of Cpr36 derivatives reacts with Malonyl CoA. (B) Mass spectrum for the protein peak eluting at time t ≈14 min of Cpr36 derivatives reacts with acetyl CoA.

The *cpr34* and *cpr35* genes were cloned and expressed in *E. coli* BL21 (DE3) as a maltose-binding protein (MBP)-fusion protein to obtain soluble protein with the expected sizes for in vitro assays. The *cpr33* gene was cloned and expressed in *E. coli* BL21 (DE3) as a pET30-*cpr33* to obtain soluble protein with the expected sizes (Figure 4.12). HPLC was used to monitor the reactions catalyzed by Cpr34/35(Figure 4.13). Interestingly, there was only a distinguished difference of elution time between apo-Cpr36 and the activated Cpr36 derivatives, which provided more evidence that the chemical moiety changes among the forms of activated Cpr36 is too small to lead to notable HPLC elution time changes.

The sequential reaction catalyzed by Cpr34/35 was characterized by using the same strategy of protein mass spectrum as before and the protein mass spectrum result of Figure 4.14 provided more evidence that Cpr34/35 catalyzed the reaction of formation of the β-ketothioester linked Cpr36 following the formation of thioester-linked PABA (expected mass, 14944; found mass, 14943).

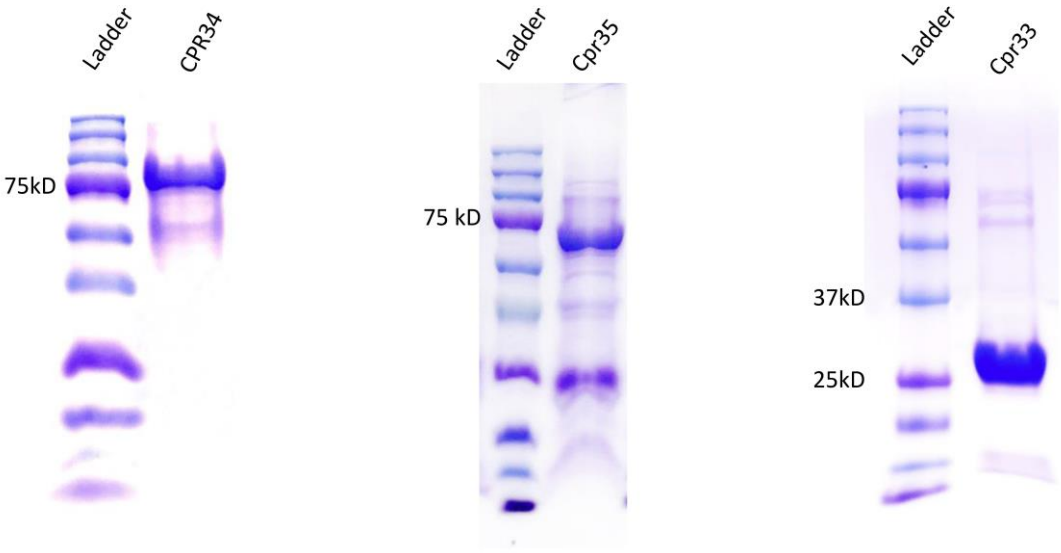

Figure 4.12 SDS-PAGE analysis of purified protein of Cpr34, Cpr35 and Cpr33 (His6-MBP-Cpr34, expected MW of 84 kD; His6-MBP-Cpr35, expected MW of 70.5 kD; Cpr33, expected MW of 30 kD).

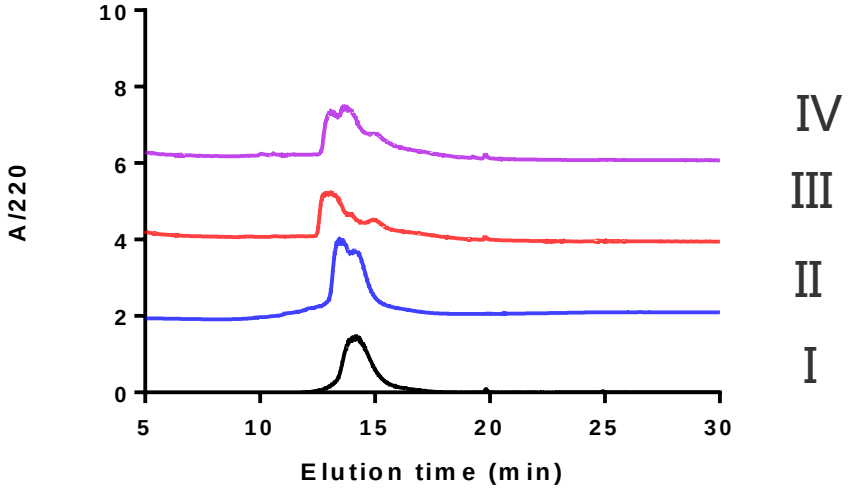

Figure 4.13 HPLC Characterization of reaction catalyzed by Cpr34/35. I, only apo-Cpr36 w/o Svp enzyme; II, 12h reaction of apo-Cpr36 with malonyl CoA catalyzing by Svp; III, 12h reaction of holo-Cpr36 with Cpr37 and PABA; IV, 12 h reaction consistent of holo-Cpr36, Cpr37, PABA and Cpr34/35.

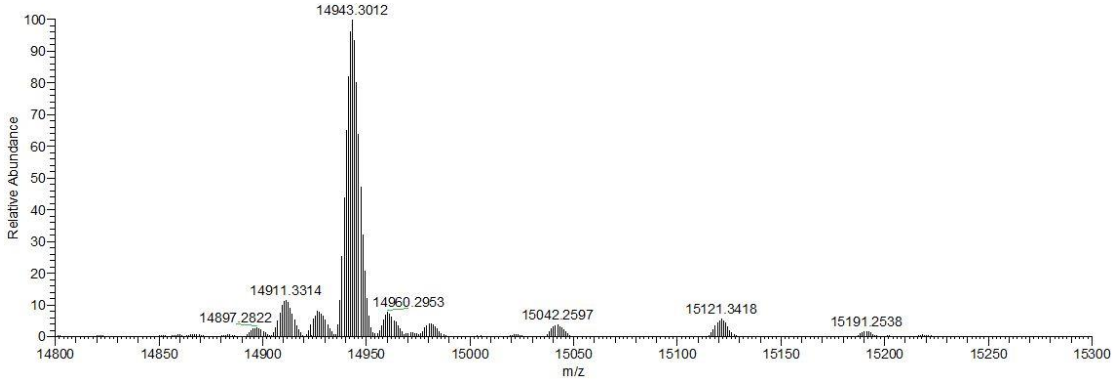

Figure 4.14 Protein mass spectrum characterization the reaction catalyzed by Cpr34/35. (β-ketothioester linked Cpr36 expected $(M+Na)^+$ ion at molecular mass of 14944 Da, found 14943 Da).

The further small disulfides molecule cleavage assay provided more evidence that decarboxylative condensation catalyzed by the Cpr34/35 heterodimer was completed and β-ketothioester linked Cpr36 was formed (Figure 4.15 and 4.16). HPLC characterization revealed that the product of reaction catalyzed by the enzyme of Cpr33 had ~0.5 minute earlier retention time comparing to the product of reaction catalyzed by the enzyme of Cpr37 and Cpr34/35 (Figure 4.15). It provided more evidence that

the reduction from ketone moiety to hydroxyl happened and the increase polarity would lead to earlier retention time of cleaved disulfides following mild treatment of extra cysteamine.

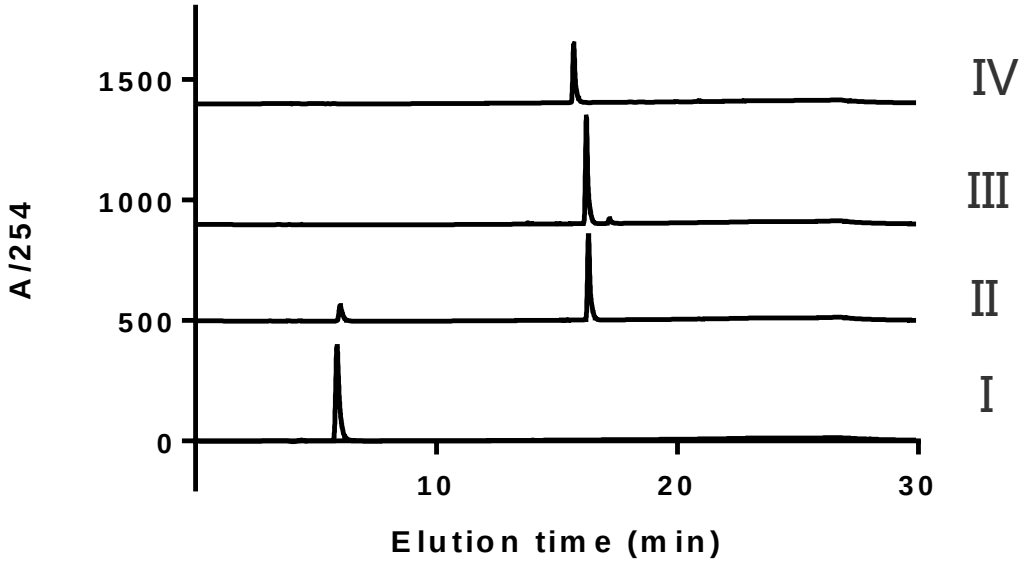

Figure 4.15 Chemical release of Cpr36-bound polyketides. Chemical release of Cpr36-bound polyketides. Derivatized polyketides released upon cysteamine treatment of purified Cpr36. HPLC traces (λ = 254 nm) of adducts extracted following treatment of a variety of Cpr36 based reaction pools with 0.25 M cysteamine are shown. I, negative control(the reaction pool of PABA, Cpr36, Svp but without Cpr37); II, the reaction pool of PABA, Cpr36, Svp and Cpr37; III, the reaction pool of PABA, Cpr36, Svp, Cpr37 and Cpr34/35; IV, the reaction pool of PABA, Cpr36, Svp, Cpr37, Cpr34/35 and Cpr33.

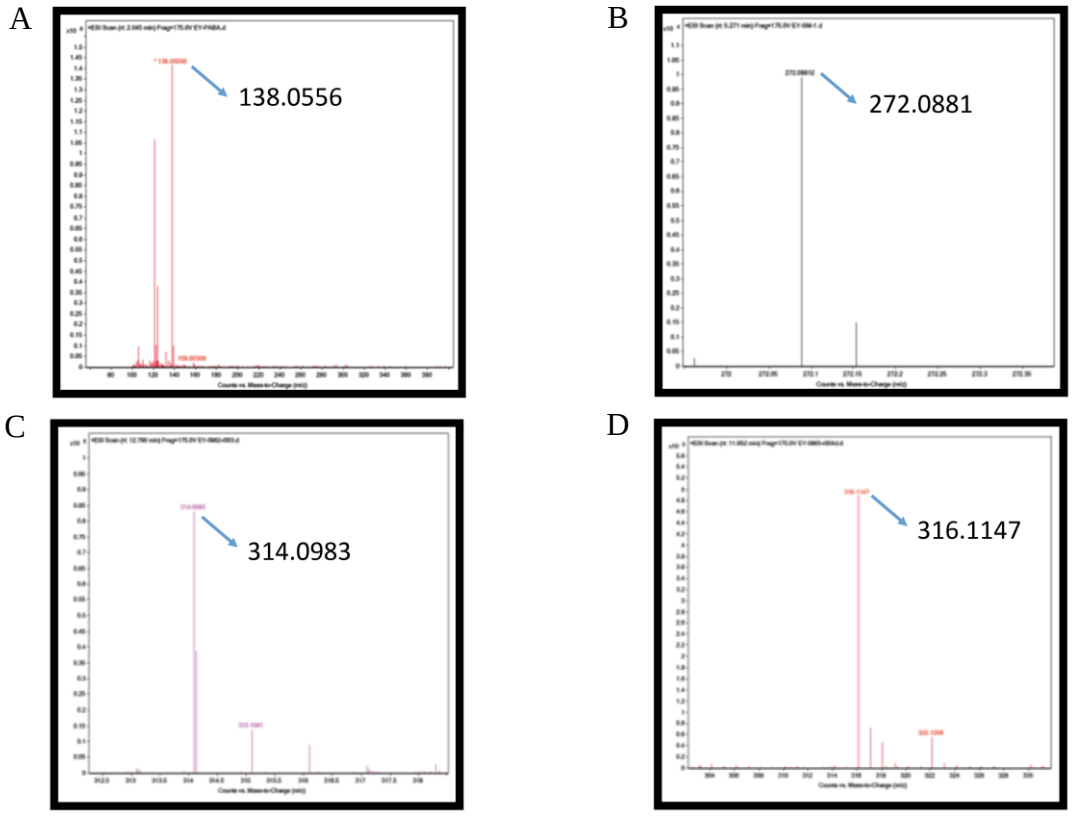

Figure 4.16 (+)-HR-ESI-MS (positive mode) of masses of Cpr36-bound polyketides **1-3a**. A, negative control (the reaction pool of PABA, Cpr36, Svp but without Cpr37); PABA expected $(M+H)^+$ ion at *m/z* =138.0550, found: 138.0556; B, the reaction pool consisted of PABA, Cpr36, Svp and Cpr37 (1a expected $(M+H)^+$ ion at *m/z* =272.0886, found:272.0881); C, the reaction pool consisted of PABA, Cpr36, Svp, Cpr37 and Cpr34/35 (2a expected $(M+H)^+$ ion at *m/z* =314.0991, found:314.0983; D, the reaction pool consisted of PABA, Cpr36, Svp, Cpr37, Cpr34/35 and Cpr33 (3a expected $(M+H)^+$ ion at *m/z* =316.1148, found: 316.1147).

The previous results of small disulfides molecule cleave assay for assaying CP-bound intermediates allow us to delineate a comprehensive picture of the chemistry of enzyme Cpr36 and downstream reaction from the perspective of both in a heterologous host and in reconstitution reactions.

To better exploit the unique structural properties of the products of these capuramycin series enzymes, especially the stereochemistry properties of product of Cpr33, reconstitution reactions in vitro conditions using the hypothetical intermediate were conducted. Therefore, **8**, the mimic of beta ketoester linked Cpr36 which was predicted

to serve as the substrate of Cpr33 was synthesized and the following reaction (Figure 4.17).

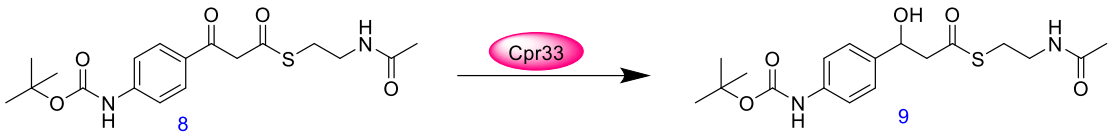

Figure 4.17 Proposed reactions catalyzed by Cpr33.

Bioinformatics analysis revealed Cpr33 has sequence similarity to proteins annotated as 3-oxoacyl-ACP reductase (FabG) from bacterial type II fatty acid synthases, an enzyme family which catalyzes hydride addition to the si face to generate the R configuration [106]. HPLC analysis of the reaction catalyzed by Cpr33 discloses there was a ~1min elution time difference between the main peak of substrate and product (Figure 4.18). Further high resolution MS analysis reveals a more convincing data with $(M + H)^+$ ion at *m/z* = 383.1642, which is consistent with the molecular formula of 9 (Figure 4.19).

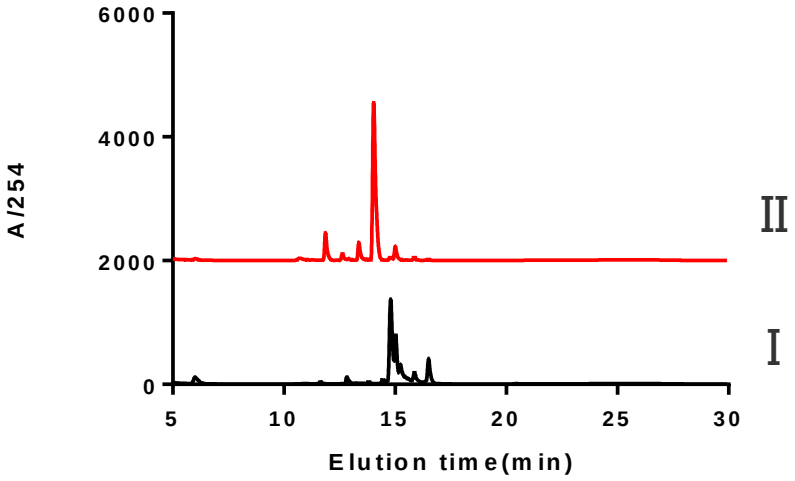

Figure 4.18 HPLC characterization of Cpr33. I, HPLC analysis of reaction with Cpr33 with NADPH 12 h; II, 12 h reaction with NADPH, w/o Cpr33.

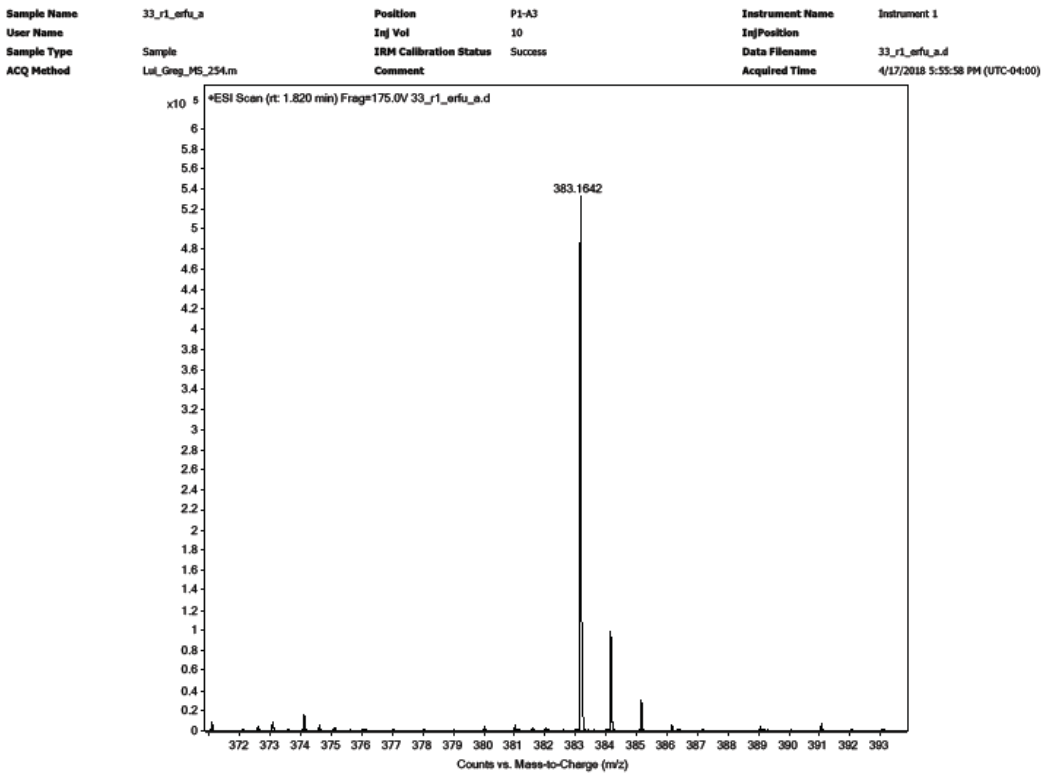

Figure 4.19 (+)-HR-ESI-MS (positive mode) of product of Cpr33 ($C_{18}H_{27}N_2O_5S$, expected $(M+H)^+$ ion at *m/z* = 383.1635, found: 383.1642).

Purified product of Cpr33 was analyzed ^1H and ^{13}C NMR spectroscopic analyses (400 MHz, $CDCl_3$) and relative peak intensities were assigned based on the natural abundance ^1H and ^{13}C NMR spectra (Figure 4.20 and Figure 4.21). Correlations obtained from COSY, NOESY, HMBC and HMQC experiments of product of Cpr33 were used to establish the structure (Table 4.3) (Figure 4.20, S.4.5-4.8).

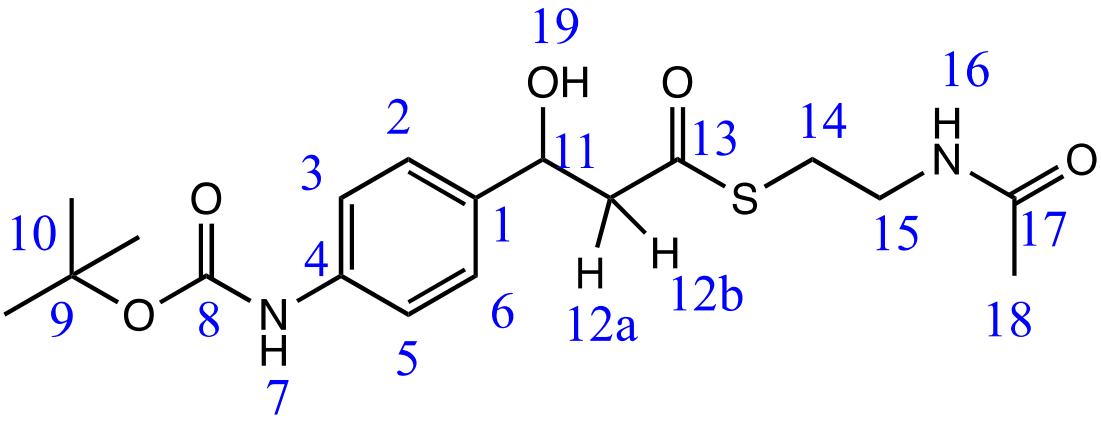

Figure 4.20 Number labeled structures of **9** for NMR.

NMR spectroscopic analyses revealed a doublet of doublets characteristic proton signal which is located in the position of C11 with chemical shift 5.14 (dd, J=8.90, 3.77 Hz, 1H) (Figure 4.21).

Table 4.3 NMR data of **9** in $CDCl_3$ (^1H: 400 MHz, ^{13}C: 100 MHz, J in Hz).

positions	δ_C	δ_H
1	137	-
2,6	118.88	7.35(d, J=8.43 Hz, 2H)
3,5	126.71	7.29(d, J=8.86 Hz, 2H)
4	138.38	-
7	-	6.48(s,1H)
8	153.16	-
9	80.98	-
10	28.61	1.52(s, 9H)
11	71.22	5.14(dd, J=8.90, 3.77 Hz, 1H)
12a	52.88	2.91(dd, J=15.32, 3.8Hz, 1H)
12b,14	52.88(12b), 29.24(14)	2.99-3.11(m, 3H)
13	199	-
15	39.5	3.36-3.49(m, 2H)
16	-	5.43 (bs, 1H)
17	170.67	-
18	23.54	1.96(s,3H)

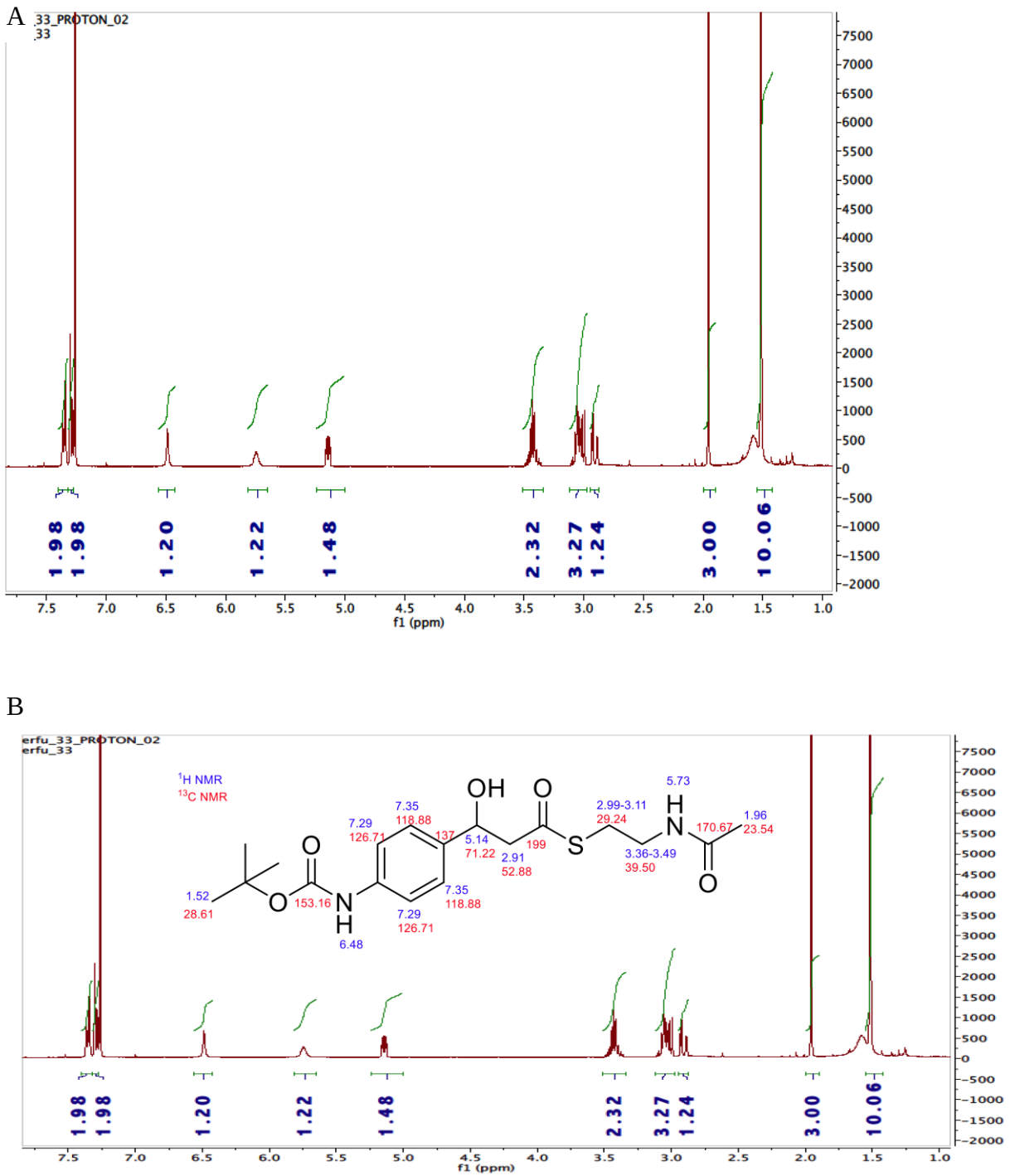

Figure 4.21 NMR spectrum of **9**. (A) ^{1}H NMR spectrum ($CDCl_3$, 400 MHz) and (B) ^{13}C NMR spectrum ($CDCl_3$, 400 MHz).

4.4 Conclusion

Enzyme Cpr34/35 heterodimer has been revealed to belong to ketosynthase family with Cpr34 as β-ketoacyl synthase and Cpr35 as α-ketoacyl synthase. The two proteins worked in concert to catalyze decarboxylative condensation between a malonyl-S-carrier protein (CP) and recipient thioester during biosynthesis of A-102395.

Cpr33 functioned as 3-oxoacyl-ACP reductase which catalyzed β-reduction to form the β-hydroxythioester intermediate. Combing with bioinformatics analysis which suggested that sequence of Cpr33 had the closest similarity to 3-oxoacyl-ACP reductase (FabG) from bacterial type II fatty acid synthases [106], it is reasonable to assume Cpr33 catalyzes hydride addition to the *re* face to generate the *S* configuration intermediate. Future effort of determining this unique stereoselective reduction may be a critical advancement for interrogating the biosynthesis of the unusual chemical components of the family of antibiotics.

4.5 Supporting Information

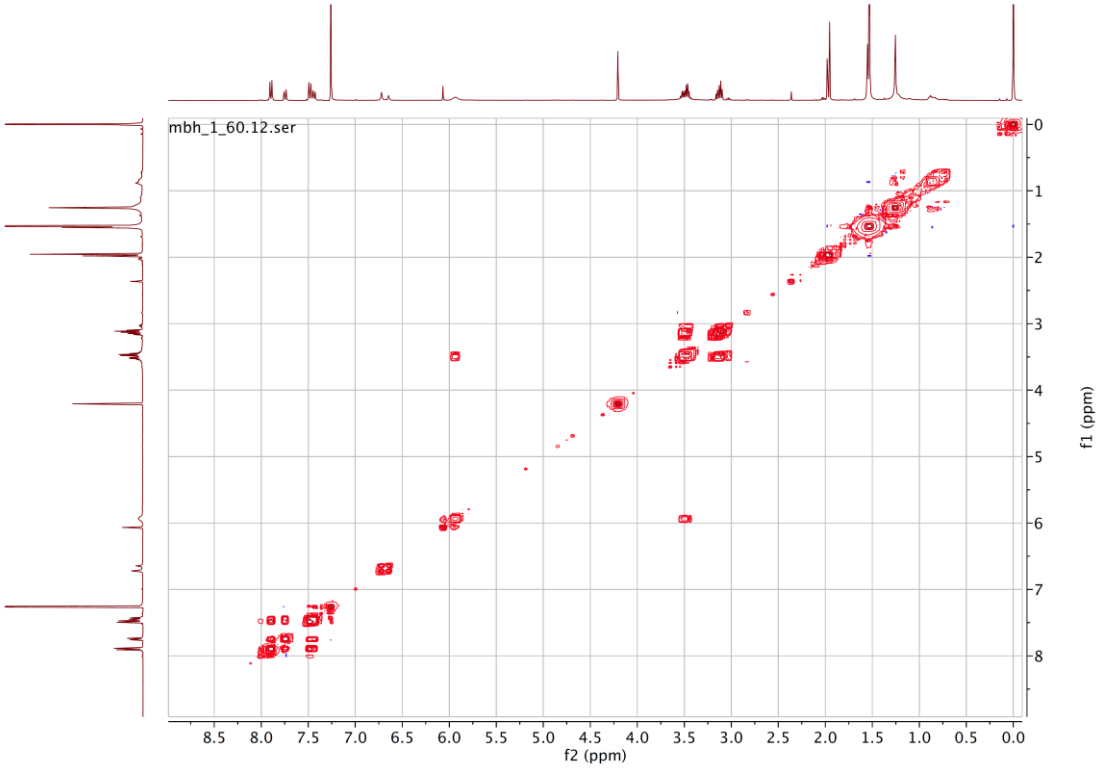

Figure S4.1 COSY spectrum (CDCl3, 400 MHz) of 8.

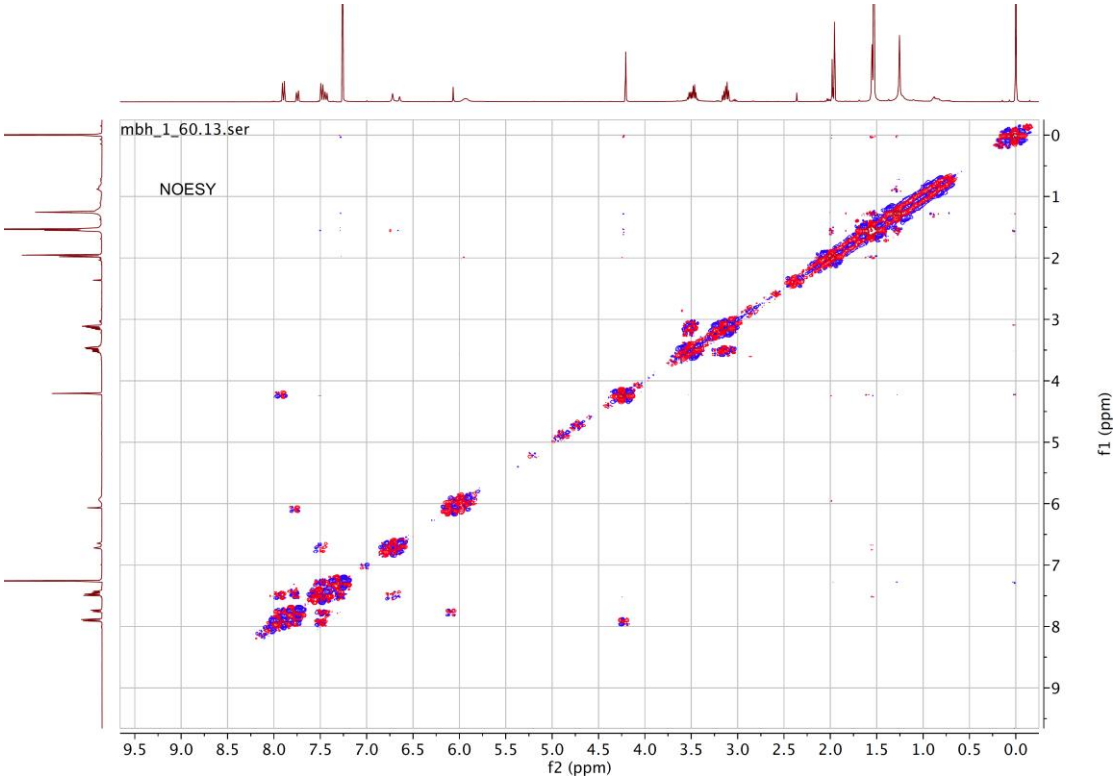

Figure S4.2 NOESY spectrum (CDCl3, 400 MHz) of 8.

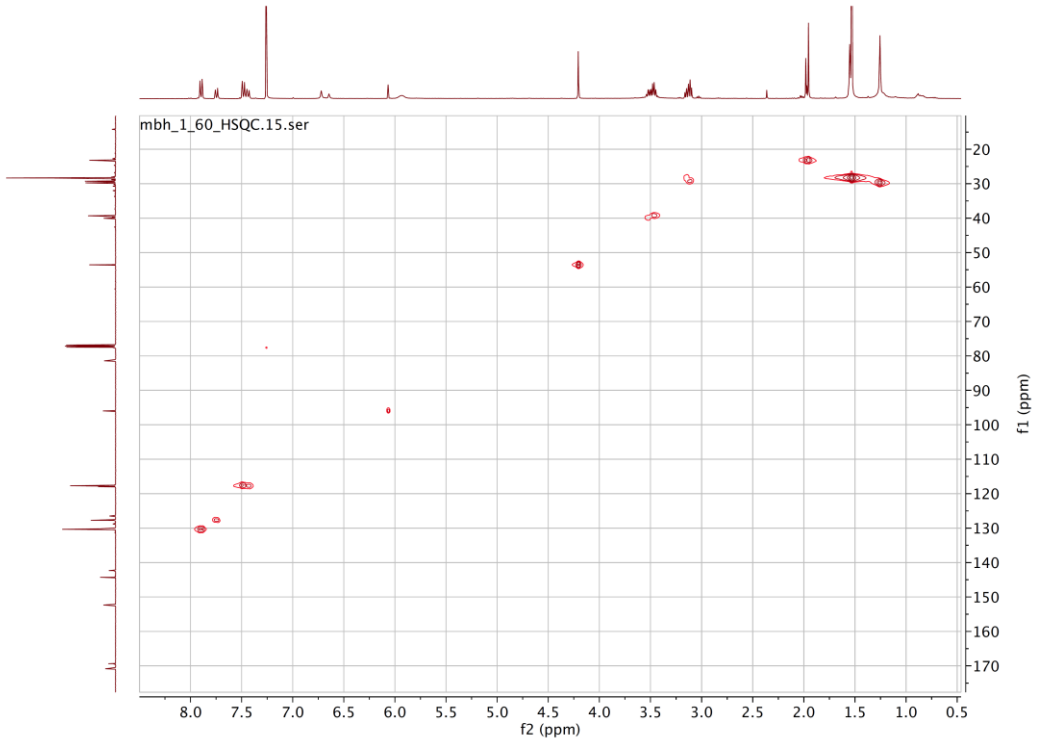

Figure S4.3 HSQC spectrum ($CDCl_3$, 400 MHz) of 8.

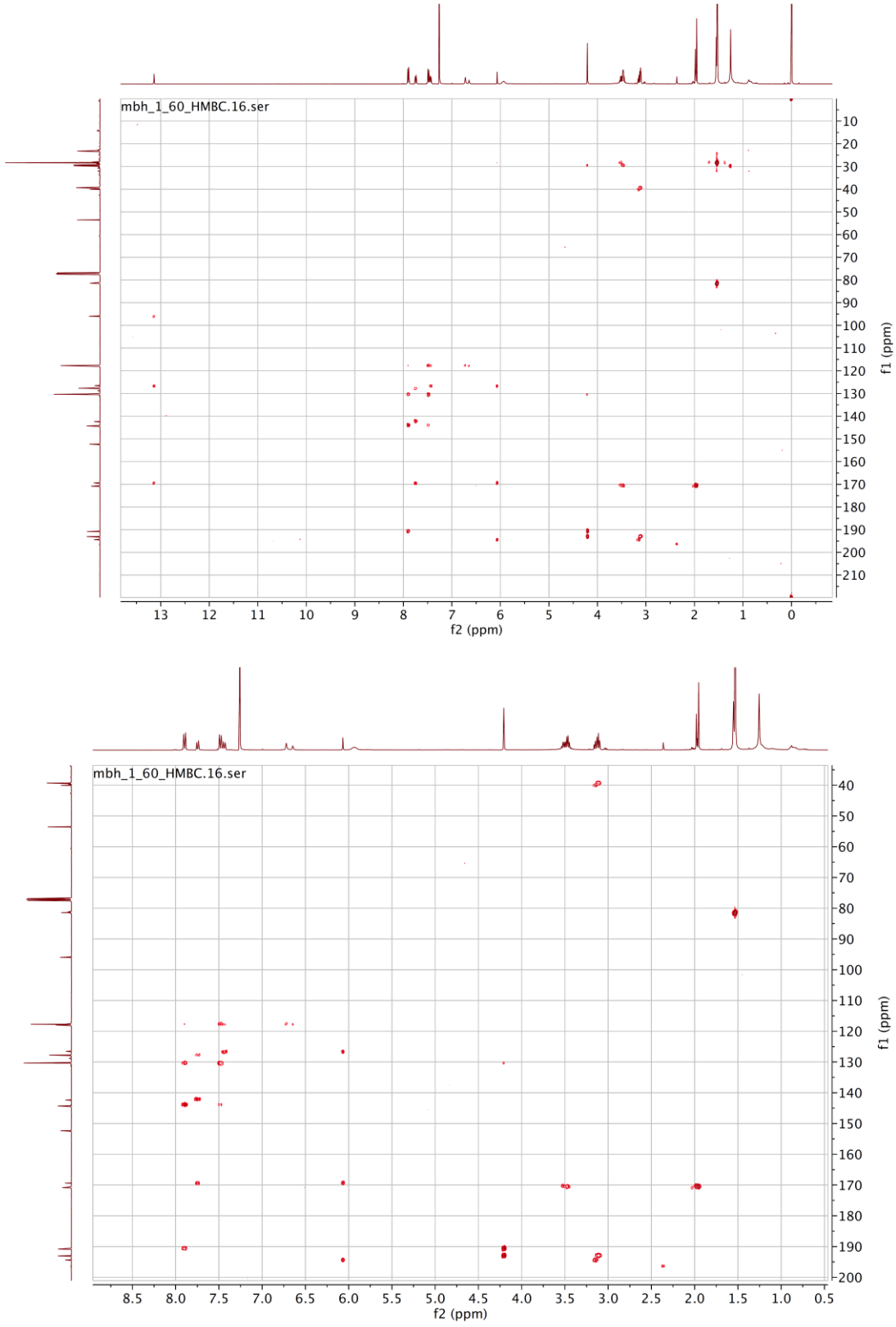

Figure S4.4 HSBC spectrum ($CDCl_3$, 400 MHz) of 8.

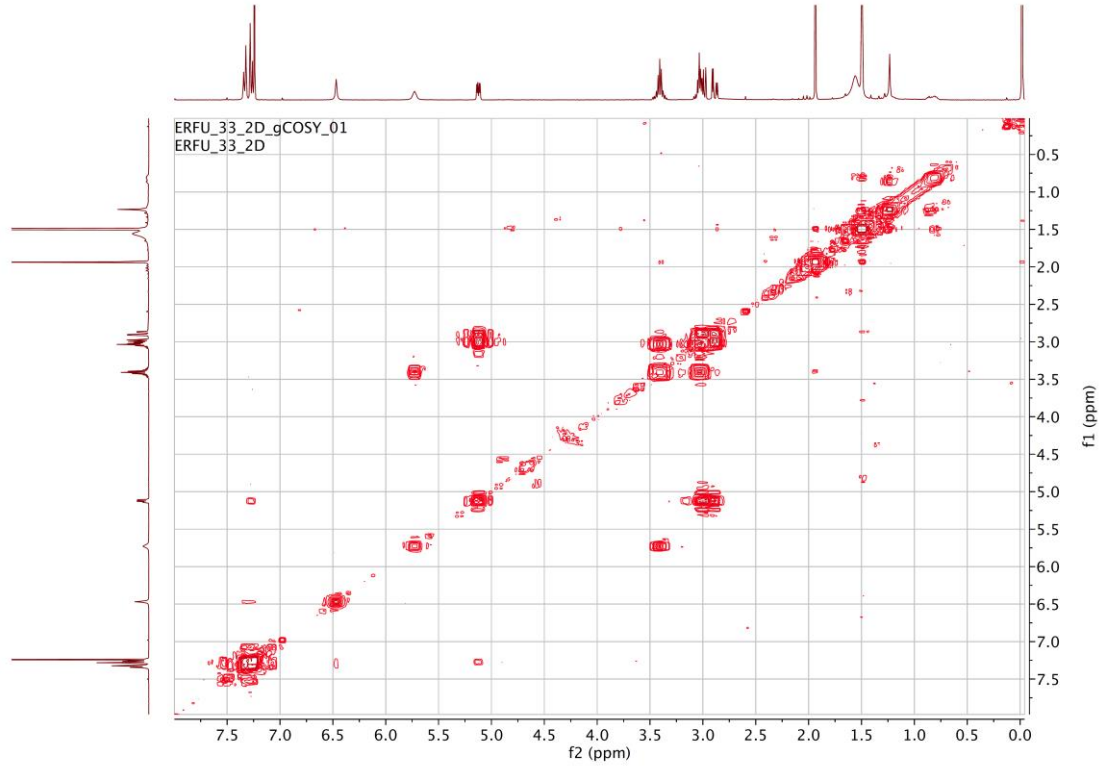

Figure S4.5 COSY spectrum ($CDCl_3$, 400 MHz) of 9.

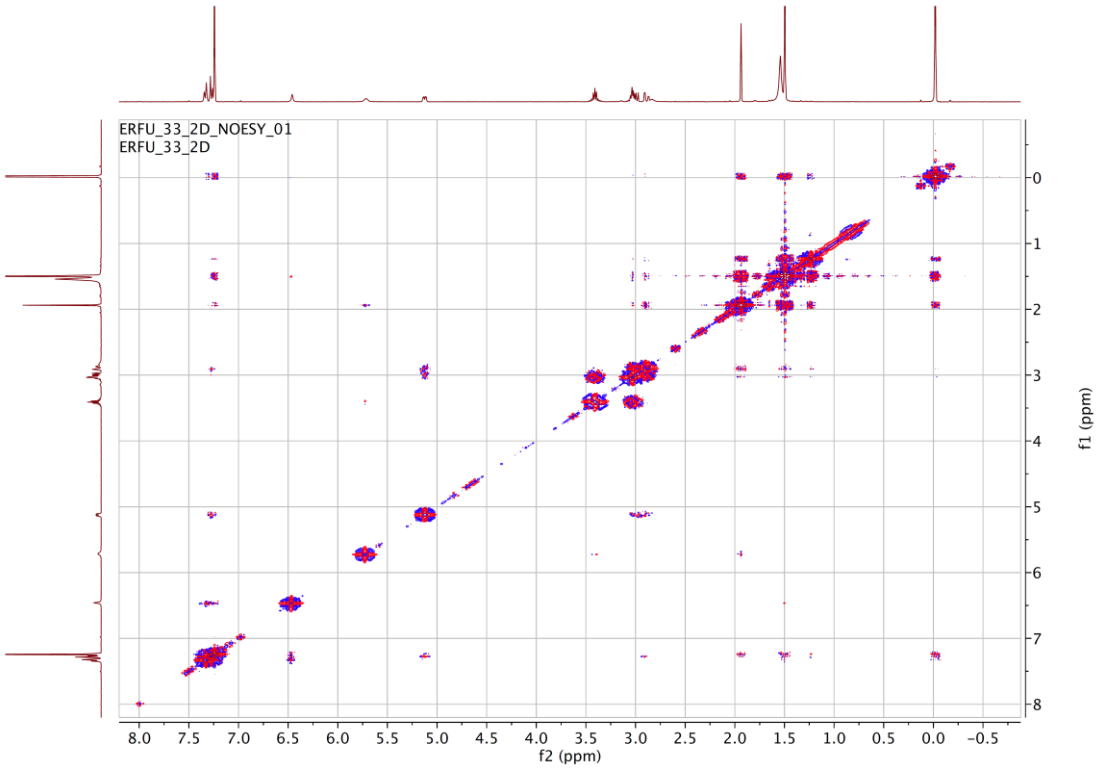

Figure S4.6 NOSY spectrum ($CDCl_3$, 400 MHz) of 9.

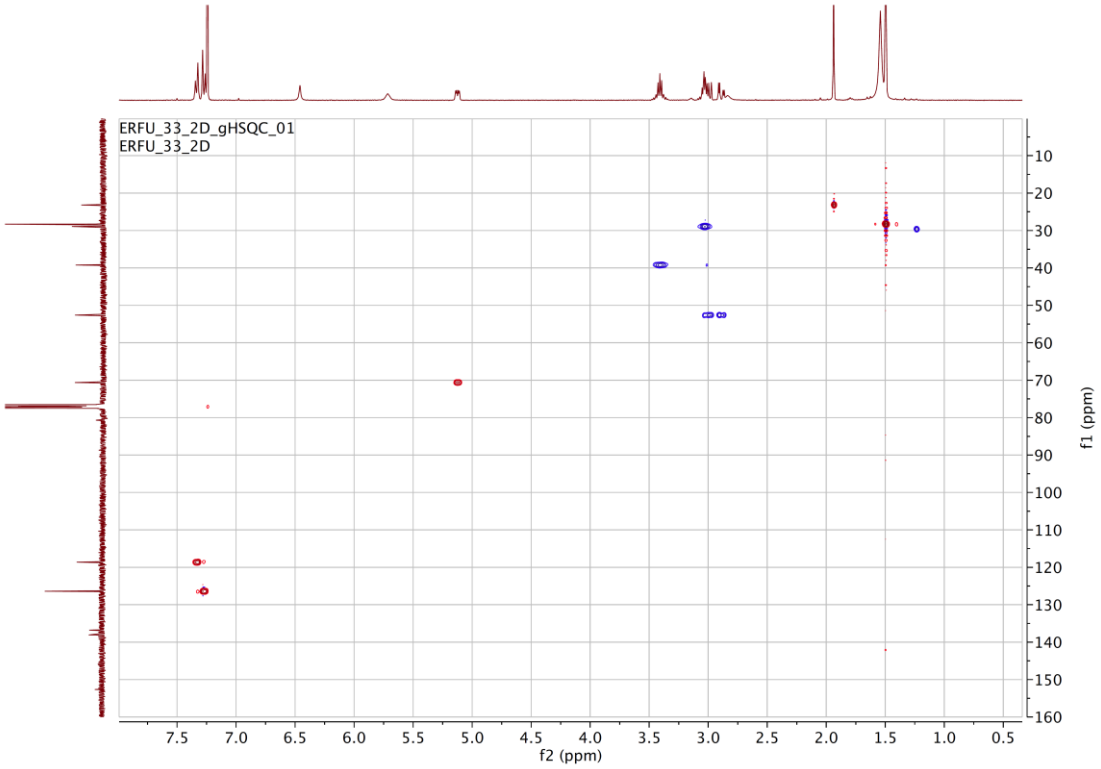

Figure S4.7 HSQCspectrum ($CDCl_3$, 400 MHz) of 9.

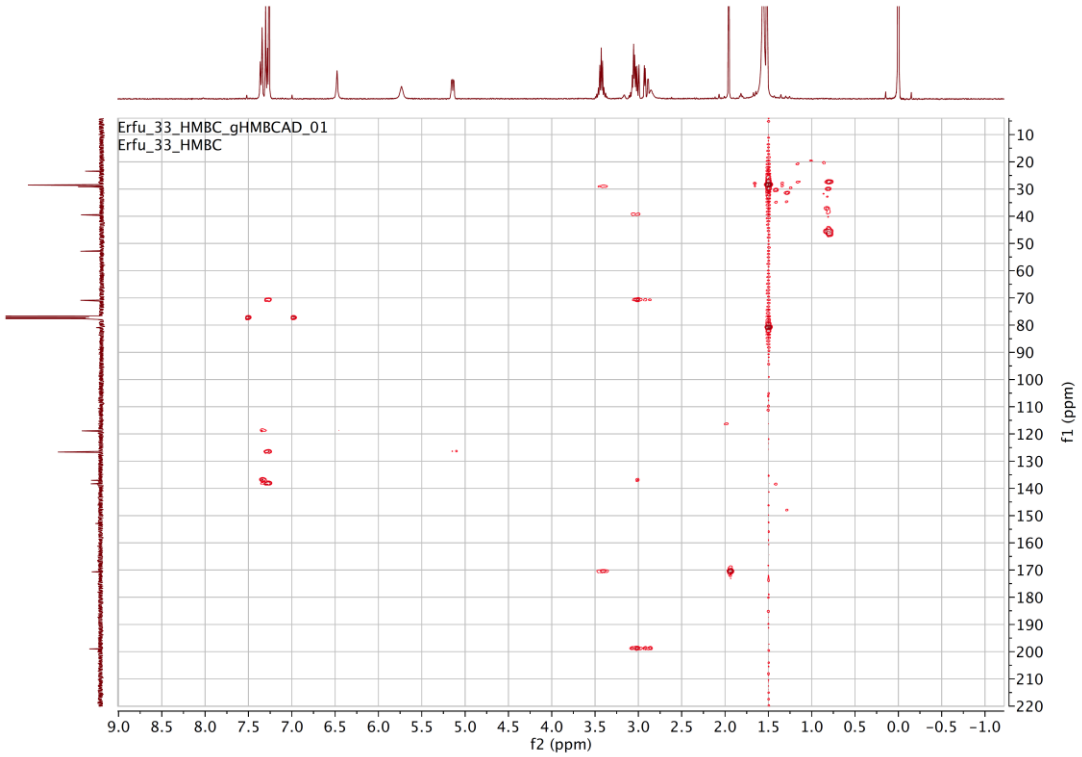

Figure S4.8 HMBCspectrum ($CDCl_3$, 400 MHz) of 9.

Chapter 5: Summary

Tuberculosis (TB) remains a significant threat to global health. However, the discovery of novel antibiotics has dramatically decreased over the last few decades. Therefore, the discovery and development of new anti-tuberculosis antibiotics are urgently needed. Capuramycin is a class of nucleoside antibiotics which show potent anti-TB activity by inhibiting the bacterial MraY which plays key role in the biosynthesis of peptidoglycan cell wall. The first part of this dissertation describes the identification and characterization of the biosynthetic gene cluster of Cpr38 and Cpr12 for A-102395 which is the latest member of capuramycins. Cpr38, a bidomain protein, catalyzes a two-step reaction involving amidohydrolysis of L-Gln with ammonia incorporated into chorismic acid to generate ADC. Cpr12, carrying similarity to proteins annotated as ADC lyase, catalyzes the elimination of pyruvate to form PABA. Cpr36, was characterized as an aryl carrier protein. Similar with other aryl carrier proteins, it can be activated catalyzing by Svp with adding 4′-PPT arm into its highly conserved serine residue. We also found that a mutation at N-63 from metheine to alaine had no effect on the aryl carrier protein function at all. The remaining steps are putatively initiated by Cpr37-catalyzed activation of PABA as the acyl-adenylate with loading to the free-standing carrier protein Cpr36 to form the thioester-linked PABA. Functional characterization of and Cpr37 indicated that it catalyzes identical reactions as AcmD and AcmA, suggesting a shared biosynthetic paradigm in antibiotics of actinomycin [201]. We have functionally and kinetically characterized Cpr37 as an adenylation domain protein with K_m= 121 ±12 μM and k_{cat}= 2.32 ±0.2min.

Cpr34 and *cpr35*, two genes adjacent to *cpr36*, have been proved to work in concert to catalyze decarboxylative condensation between a thioester linked PABA and malonyl-S-acyl carrier protein (ACP) during aromatic polyketide biosynthesis catalyzed by type II polyketide synthases. Bioinformatic analysis reveals that Cpr34 may act as β-ketoacyl-acyl-carrier-protein synthase II and Cpr35 works as α-ketoacyl-acyl-carrier-

protein synthase II.

Following condensation, Cpr33 functions as 3-oxoacyl-ACP reductase to catalyze reduction to the β-hydroxythioester intermediate. Bioinformatics analysis suggests that sequence of Cpr33 has the closest similarity to 3-oxoacyl-ACP reductase (FabG) from bacterial type II fatty acid synthases, which has been proved to catalyze hydride addition to the *si* face to generate the *R* configuration (71). It is thus reasonable to assume that Cpr33 catalyze the same reduction reaction in which hydride addition to the *re* face to generate the *S* configuration. However, the β-hydroxythioester intermediate stereochemistry in the biosynthesis pathway of A-102395 still need to be established.

The remaining steps which include α-hydroxylation and formation of two amide bonds in polyamide biosynthesis is expected to be catalyzed putative proteins encoded within ORFs *cpr47–cpr57*. These proteins include but not limited to an adenylation-like domain protein with predicted specificity to a hydrophilic amino acid (Cpr54), two carrier proteins (Cpr48 and Cpr55), a condensation domain protein (Cpr47), and three transglutaminase-like proteins (Cpr49, Cpr50, and Cpr57). Further study are still needed to illustrate the profound machinery for interrogating the biosynthesis of the unusual chemical components of this family of antibiotics.

References

1. Koehn, F.E. and G.T. Carter, *The evolving role of natural products in drug discovery.* Nature Reviews Drug Discovery, 2005. **4**: p. 206.
2. Harvey, A.L., *Natural products in drug discovery.* Drug Discovery Today, 2008. **13**(19): p. 894-901.
3. Butler, M.S., *Natural products to drugs: natural product derived compounds in clinical trials.* Natural Product Reports, 2005. **22**(2): p. 162-195.
4. Li, J.W.-H. and J.C. Vederas, *Drug Discovery and Natural Products: End of an Era or an Endless Frontier?* Science, 2009. **325**(5937): p. 161-165.
5. Davies, H.M.L., *Synthetic lessons from nature.* Nature, 2009. **459**: p. 786.
6. Walsh, C.T. and M.A. Fischbach, *Natural Products Version 2.0: Connecting Genes to Molecules.* Journal of the American Chemical Society, 2010. **132**(8): p. 2469-2493.
7. Watve, M.G., et al., *How many antibiotics are produced by the genus Streptomyces?* Archives of Microbiology, 2001. **176**(5): p. 386-390.
8. Baltz, R.H., *Marcel Faber Roundtable: Is our antibiotic pipeline unproductive because of starvation, constipation or lack of inspiration?* Journal of Industrial Microbiology and Biotechnology, 2006. **33**(7): p. 507-513.
9. Newman, D.J. and G.M. Cragg, *Natural Products As Sources of New Drugs over the 30 Years from 1981 to 2010.* Journal of Natural Products, 2012. **75**(3): p. 311-335.
10. Bush, K., *Improving known classes of antibiotics: an optimistic approach for the future.* Current Opinion in Pharmacology, 2012. **12**(5): p. 527-534.
11. O'Connell, K.M.G., et al., *Combating Multidrug-Resistant Bacteria: Current Strategies for the Discovery of Novel Antibacterials.* Angewandte Chemie International Edition, 2013. **52**(41): p. 10706-10733.
12. Hampton, T., *Novel Programs and Discoveries Aim to Combat Antibiotic Resistance.* Jama, 2015. **313**(24): p. 2411-3.
13. Brennan, P.J., *Structure, function, and biogenesis of the cell wall of Mycobacterium tuberculosis.* Tuberculosis. **83**(1): p. 91-97.
14. Takayama, K., C. Wang, and G.S. Besra, *Pathway to Synthesis and Processing of Mycolic Acids in Mycobacterium tuberculosis.* Clinical Microbiology Reviews, 2005. **18**(1): p. 81-101.
15. Mase, S., et al., *Provisional CDC Guidelines for the Use and Safety Monitoring of Bedaquiline Fumarate (Sirturo) for the Treatment of Multidrug-Resistant Tuberculosis.* Morbidity and Mortality Weekly Report: Recommendations and Reports, 2013. **62**(9): p. 1-12.
16. Spížek, J., et al., *Do we need new antibiotics? The search for new targets and new compounds.* Journal of Industrial Microbiology & Biotechnology, 2010. **37**(12): p. 1241-1248.
17. WHO, *Global Tuberculosis Report 2014.* World Health Organization, 2014.
18. Friedman, D., and Alper, J., *Technological Challenges in Antibiotic Discovery and Development:: A Workshop Summary,.* National Academies Press, 2014.
19. Lewis, K., *Platforms for antibiotic discovery.* Nat Rev Drug Discov, 2013. **12**(5): p. 371-87.
20. Koul, A., et al., *Diarylquinolines target subunit c of mycobacterial ATP synthase.* Nature Chemical Biology, 2007. **3**: p. 323.

21. Lu, P., H. Lill, and D. Bald, *ATP synthase in mycobacteria: Special features and implications for a function as drug target.* Biochimica et Biophysica Acta (BBA) - Bioenergetics, 2014. **1837**(7): p. 1208-1218.
22. Bélard, S., et al., *Bedaquiline for the treatment of drug-resistant tuberculosis.* Expert Review of Anti-infective Therapy, 2015. **13**(5): p. 535-553.
23. T. D. H. Bugg, C.T.W., *Intracellular steps of bacterial cell wall peptidoglycan biosynthesis: enzymology, antibiotics, and antibiotic resistance.* Natural Product Reports, 1992. **9**(3): p. 199-215.
24. Park, J.T. and M.J. Johnson, *ACCUMULATION OF LABILE PHOSPHATE IN STAPHYLOCOCCUS AUREUS GROWN IN THE PRESENCE OF PENICILLIN.* Journal of Biological Chemistry, 1949. **179**(2): p. 585-592.
25. Heijenoort, J.v., *Recent advances in the formation of the bacterial peptidoglycan monomer unit.* Natural Product Reports, 2001. **18**(5): p. 503-519.
26. Bugg, T.D.H., *Comprehensive Natural Products Chemistry.* 1999. **3**: p. 241–294.
27. Winn, M., et al., *Antimicrobial nucleoside antibiotics targeting cell wall assembly: Recent advances in structure-function studies and nucleoside biosynthesis.* Natural Product Reports, 2010. **27**(2): p. 279-304.
28. Kimura, K.-i. and T.D.H. Bugg, *Recent advances in antimicrobial nucleoside antibiotics targeting cell wall biosynthesis.* Natural Product Reports, 2003. **20**(2): p. 252-273.
29. Dini, C., *MraY Inhibitors as Novel Antibacterial Agents.* Current Topics in Medicinal Chemistry, 2005. **5**: p. 1221-1236.
30. Bugg, T.D.H., *3.10 - Bacterial Peptidoglycan Biosynthesis and its Inhibition, in: S.D.B.N. Meth-Cohn (Ed.).* Comprehensive Natural Products Chemistry, 1999: p. 241-294.
31. Bouhss, A., et al., *The biosynthesis of peptidoglycan lipid-linked intermediates.* FEMS Microbiology Reviews, 2008. **32**(2): p. 208-233.
32. Sauvage, E., et al., *The penicillin-binding proteins: structure and role in peptidoglycan biosynthesis.* FEMS Microbiology Reviews, 2008. **32**(2): p. 234-258.
33. Walker, S., et al., *Chemistry and Biology of Ramoplanin: A Lipoglycodepsipeptide with Potent Antibiotic Activity.* Chemical Reviews, 2005. **105**(2): p. 449-476.
34. Kahne, D., et al., *Glycopeptide and Lipoglycopeptide Antibiotics.* Chemical Reviews, 2005. **105**(2): p. 425-448.
35. Böth, D., G. Schneider, and R. Schnell, *Peptidoglycan Remodeling in Mycobacterium tuberculosis: Comparison of Structures and Catalytic Activities of RipA and RipB.* Journal of Molecular Biology, 2011. **413**(1): p. 247-260.
36. Ikeda, M., et al., *The Escherichia coli mraY gene encoding UDP-N-acetylmuramoyl-pentapeptide: undecaprenyl-phosphate phospho-N-acetylmuramoyl-pentapeptide transferase.* Journal of Bacteriology, 1991. **173**(3): p. 1021-1026.
37. Pless, D.D. and F.C. Neuhaus, *Initial Membrane Reaction in Peptidoglycan Synthesis: LIPID DEPENDENCE OF PHOSPHO-N-ACETYLMURAMYL-PENTAPEPTIDE TRANSLOCASE (EXCHANGE REACTION).* Journal of Biological Chemistry, 1973. **248**(5): p. 1568-1576.
38. Lehrman, M.A., *Commentary: A family of UDP-GlcNAc/MurNAc: polyisoprenol-P GlcNAc/MurNAc-1-P transferases.* Glycobiology, 1994. **4**(6): p. 768-771.

39. Boyle, D.S. and W.D. Donachie, *mraY Is an Essential Gene for Cell Growth in Escherichia coli.* Journal of Bacteriology, 1998. **180**(23): p. 6429-6432.
40. Chung, B.C., et al., *Crystal Structure of MraY, an Essential Membrane Enzyme for Bacterial Cell Wall Synthesis.* Science, 2013. **341**(6149): p. 1012-1016.
41. Geis, A. and R. Plapp, *Phospho-N-acetylmuramoyl-pentapeptide-transferase of Escherichia coli K12 Properties of the membrane-bound and the extracted and partially purified enzyme.* Biochimica et Biophysica Acta (BBA) - Enzymology, 1978. **527**(2): p. 414-424.
42. Bouhss, A., et al., *Topological analysis of the MraY protein catalysing the first membrane step of peptidoglycan synthesis.* Molecular Microbiology, 1999. **34**(3): p. 576-585.
43. Lloyd, A.J., et al., *Phospho-N-Acetyl-Muramyl-Pentapeptide Translocase from Escherichia coli: Catalytic Role of Conserved Aspartic Acid Residues.* Journal of Bacteriology, 2004. **186**(6): p. 1747-1757.
44. Muramatsu, Y., et al., *Studies on Novel Bacterial Translocase I Inhibitors, A-500359s.* The Journal Of Antibiotics, 2006. **59**: p. 601.
45. Bernhardt, T.G., W.D. Roof, and R. Young, *Genetic evidence that the bacteriophage φX174 lysis protein inhibits cell wall synthesis.* Proceedings of the National Academy of Sciences, 2000. **97**(8): p. 4297-4302.
46. Bernhardt, T.G., D.K. Struck, and R. Young, *The Lysis Protein E of φX174 Is a Specific Inhibitor of the MraY-catalyzed Step in Peptidoglycan Synthesis.* Journal of Biological Chemistry, 2001. **276**(9): p. 6093-6097.
47. Tanaka, S. and W.M. Clemons, *Minimal requirements for inhibition of MraY by lysis protein E from bacteriophage ΦX174.* Molecular microbiology, 2012. **85**(5): p. 975-985.
48. Mendel, S., et al., *Interaction of the transmembrane domain of lysis protein E from bacteriophage ϕX174 with bacterial translocase MraY and peptidyl-prolyl isomerase SlyD.* Microbiology, 2006. **152**(10): p. 2959-2967.
49. Tanaka, H., et al., *Amphomycin inhibits phospho-N-acetylmuramyl-pentapeptide translocase in peptidoglycan synthesis of Bacillus.* Biochemical and Biophysical Research Communications, 1979. **86**(3): p. 902-908.
50. Banerjee, D.K., *Amphomycin inhibits mannosylphosphoryldolichol synthesis by forming a complex with dolichylmonophosphate.* Journal of Biological Chemistry, 1989. **264**(4): p. 2024-8.
51. Kang, M.S., J.P. Spencer, and A.D. Elbein, *Amphomycin inhibits the incorporation of mannose and GlcNAc into lipid-linked saccharides by aorta extracts.* Biochemical and Biophysical Research Communications, 1978. **82**(2): p. 568-574.
52. Kang, M.S., J.P. Spencer, and A.D. Elbein, *Amphomycin inhibition of mannose and GlcNAc incorporation into lipid-linked saccharides.* Journal of Biological Chemistry, 1978. **253**(24): p. 8860-8866.
53. B.A. Chen RH, W.D., McAlpine JB., *Pacidamycins, a novel series of antibiotics with anti-Pseudomonas aeruginosa activity. II. Isolation and structural elucidation.* The Journal of Antibiotics (Tokyo), 1989. **42**: p. 512-612.
54. Isono, F. and M. Inukai, *Mureidomycin A, a new inhibitor of bacterial peptidoglycan synthesis.* Antimicrobial Agents and Chemotherapy, 1991. **35**(2): p. 234-236.

55. Tamura, G., et al., *Tunicamycin Inhibits the Formation of Lipid Intermediate in Cell-free Peptidoglycan Synthesis of Bacteria.* Agricultural and Biological Chemistry, 1976. **40**(2): p. 447-449.
56. Muramatsu, Y., et al., *A-503083 A, B, E and F, novel inhibitors of bacterial translocase I, produced by Streptomyces sp. SANK 62799.* J Antibiot (Tokyo), 2004. **57**(10): p. 639-46.
57. Muramatsu, Y., et al., *Studies on novel bacterial translocase I inhibitors, A-500359s. I. Taxonomy, fermentation, isolation, physico-chemical properties and structure elucidation of A-500359 A, C, D and G.* J Antibiot (Tokyo), 2003. **56**(3): p. 243-52.
58. Murakami, R., et al., *A-102395, a new inhibitor of bacterial translocase I, produced by Amycolatopsis sp. SANK 60206.* J Antibiot (Tokyo), 2007. **60**(11): p. 690-5.
59. Funabashi, M., et al., *The Biosynthesis of Liposidomycin-like A-90289 Antibiotics Featuring a New Type of Sulfotransferase.* ChemBioChem, 2010. **11**(2): p. 184-190.
60. Ochi, K.E., M.; Iwani, M.; Komori, T.; Kohsaka, M. Eur., *Patent 333177 A2.* 1989.
61. Igarashi, M., et al., *Caprazamycins, Novel Lipo-nucleoside Antibiotics, from Streptomyces sp.* The Journal Of Antibiotics, 2005. **58**: p. 327.
62. Igarashi, M., et al., *Caprazamycins, novel lipo-nucleoside antibiotics, from Streptomyces sp. II. Structure elucidation of caprazamycins.* J Antibiot (Tokyo), 2005. **58**(5): p. 327-37.
63. Inukai, M., F. Isono, and A. Takatsuki, *Selective inhibition of the bacterial translocase reaction in peptidoglycan synthesis by mureidomycins.* Antimicrob Agents Chemother, 1993. **37**(5): p. 980-3.
64. Elbein, A.D., *The tunicamycins — useful tools for studies on glycoproteins.* Trends in Biochemical Sciences, 1981. **6**: p. 219-221.
65. Heifetz, A., R.W. Keenan, and A.D. Elbein, *Mechanism of action of tunicamycin on the UDP-GlcNAc:dolichyl-phosphate GlcNAc-1-phosphate transferase.* Biochemistry, 1979. **18**(11): p. 2186-2192.
66. Hirano, S., S. Ichikawa, and A. Matsuda, *Structure–activity relationship of truncated analogs of caprazamycins as potential anti-tuberculosis agents.* Bioorganic & Medicinal Chemistry, 2008. **16**(9): p. 5123-5133.
67. Dini, C., et al., *Synthesis of sub-micromolar inhibitors of MraY by exploring the region originally occupied by the diazepanone ring in the liposidomycin structure.* Bioorganic & Medicinal Chemistry Letters, 2002. **12**(8): p. 1209-1213.
68. Muramatsu, Y., M.M. Ishii, and M. Inukai, *Studies on novel bacterial translocase I inhibitors, A-500359s. II. Biological activities of A-500359 A, C, D and G.* J Antibiot (Tokyo), 2003. **56**(3): p. 253-8.
69. Brandish, P.E., et al., *Modes of action of tunicamycin, liposidomycin B, and mureidomycin A: inhibition of phospho-N-acetylmuramyl-pentapeptide translocase from Escherichia coli.* Antimicrob Agents Chemother, 1996. **40**(7): p. 1640-4.
70. Strieker, M., A. Tanovic, and M.A. Marahiel, *Nonribosomal peptide synthetases: structures and dynamics.* Curr Opin Struct Biol, 2010. **20**(2): p. 234-40.
71. Weckermann, R., R. Furbass, and M.A. Marahiel, *Complete nucleotide sequence of the tycA gene coding the tyrocidine synthetase 1 from Bacillus brevis.* Nucleic Acids Res, 1988. **16**(24): p. 11841.

72. Kratzschmar, J., M. Krause, and M.A. Marahiel, *Gramicidin S biosynthesis operon containing the structural genes grsA and grsB has an open reading frame encoding a protein homologous to fatty acid thioesterases.* J Bacteriol, 1989. **171**(10): p. 5422-9.
73. Mittenhuber, G., R. Weckermann, and M.A. Marahiel, *Gene cluster containing the genes for tyrocidine synthetases 1 and 2 from Bacillus brevis: evidence for an operon.* J Bacteriol, 1989. **171**(9): p. 4881-7.
74. Calcott, M.J. and D.F. Ackerley, *Genetic manipulation of non-ribosomal peptide synthetases to generate novel bioactive peptide products.* Biotechnol Lett, 2014. **36**(12): p. 2407-16.
75. Condurso, H.L. and S.D. Bruner, *Structure guided approaches toward exploiting and manipulating nonribosomal peptide and polyketide biosynthetic pathways.* Current Opinion in Chemical Biology, 2012. **16**(1): p. 162-169.
76. Spæren, U., L.O. Frøholm, and S.G. Laland, *Further studies on the biosynthesis of gramicidin S and proteins in a cell-free system from Bacillus brevis.* Biochemical Journal, 1967. **102**(2): p. 586-592.
77. Quadri, L.E.N., et al., *Characterization of Sfp, a Bacillus subtilis Phosphopantetheinyl Transferase for Peptidyl Carrier Protein Domains in Peptide Synthetases.* Biochemistry, 1998. **37**(6): p. 1585-1595.
78. Meluzzi, D., et al., *Top-down mass spectrometry on low-resolution instruments: Characterization of phosphopantetheinylated carrier domains in polyketide and non-ribosomal biosynthetic pathways.* Bioorganic & Medicinal Chemistry Letters, 2008. **18**(10): p. 3107-3111.
79. Fischbach, M.A., C.T. Walsh, and J. Clardy, *The evolution of gene collectives: How natural selection drives chemical innovation.* Proceedings of the National Academy of Sciences, 2008. **105**(12): p. 4601-4608.
80. Domon, B. and R. Aebersold, *Mass spectrometry and protein analysis.* Science, 2006. **312**(5771): p. 212-7.
81. Chen, H., et al., *Epothilone biosynthesis: assembly of the methylthiazolylcarboxy starter unit on the EpoB subunit.* Chem Biol, 2001. **8**(9): p. 899-912.
82. Hu, Q., et al., *The Orbitrap: a new mass spectrometer.* Journal of Mass Spectrometry, 2005. **40**(4): p. 430-443.
83. Makarov, A., *Electrostatic Axially Harmonic Orbital Trapping: A High-Performance Technique of Mass Analysis.* Analytical Chemistry, 2000. **72**(6): p. 1156-1162.
84. Makarov, A., *Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis.* Analytical chemistry, 2000. **72**(6): p. 1156-1162.
85. Gross, J., *Mass Spectrometry.* Springer Berlin Heidelberg, 2011: p. p 21-66.
86. Xian, F., C.L. Hendrickson, and A.G. Marshall, *High Resolution Mass Spectrometry.* Analytical Chemistry, 2012. **84**(2): p. 708-719.
87. Michalski, A., et al., *Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer.* Molecular & Cellular Proteomics, 2011.
88. Marahiel, M.A., T. Stachelhaus, and H.D. Mootz, *Modular Peptide Synthetases Involved in Nonribosomal Peptide Synthesis.* Chemical Reviews, 1997. **97**(7): p. 2651-2674.

89. Weber, G., et al., *The peptide synthetase catalyzing cyclosporine production in Tolypocladium niveum is encoded by a giant 45.8-kilobase open reading frame.* Current Genetics, 1994. **26**(2): p. 120-125.
90. Vanholme, R., et al., *Lignin Biosynthesis and Structure.* Plant Physiology, 2010. **153**(3): p. 895-905.
91. Reale, S., et al., *Mass spectrometry in the biosynthetic and structural investigation of lignins.* Mass Spectrom Rev, 2004. **23**(2): p. 87-126.
92. Yoshioka, K., D. Ando, and T. Watanabe, *A comparative study of matrix- and nano-assisted laser desorption/ionisation time-of-flight mass spectrometry of isolated and synthetic lignin.* Phytochem Anal, 2012. **23**(3): p. 248-53.
93. Yamaguchi, H., et al., *Capuramycin, a new nucleoside antibiotic. Taxonomy, fermentation, isolation and characterization.* J Antibiot (Tokyo), 1986. **39**(8): p. 1047-53.
94. Ohnuki, T., et al., *Studies on novel bacterial translocase I inhibitors, A-500359s. IV. Biosynthesis of A-500359s.* J Antibiot (Tokyo), 2003. **56**(3): p. 268-79.
95. Hartkoorn, R.C., et al., *Towards a new tuberculosis drug: pyridomycin – nature's isoniazid.* EMBO Molecular Medicine, 2012. **4**(10): p. 1032-1042.
96. Wang, Y., et al., *Improved synthesis of capuramycin and its analogues.* Chemistry, 2013. **19**(41): p. 13847-58.
97. Nikonenko, B.V., et al., *Activity of SQ641, a Capuramycin Analog, in a Murine Model of Tuberculosis.* Antimicrobial Agents and Chemotherapy, 2009. **53**(7): p. 3138-3139.
98. Koga, T., et al., *Activity of capuramycin analogues against Mycobacterium tuberculosis, Mycobacterium avium and Mycobacterium intracellularein vitro and in vivo.* Journal of Antimicrobial Chemotherapy, 2004. **54**(4): p. 755-760.
99. Bogatcheva, E., et al., *Chemical modification of capuramycins to enhance antibacterial activity.* Journal of Antimicrobial Chemotherapy, 2011. **66**(3): p. 578-587.
100. Reddy, V.M., L. Einck, and C.A. Nacy, *In Vitro Antimycobacterial Activities of Capuramycin Analogues.* Antimicrobial Agents and Chemotherapy, 2008. **52**(2): p. 719-721.
101. Funabashi, M., et al., *An ATP-independent strategy for amide bond formation in antibiotic biosynthesis.* Nature Chemical Biology, 2010. **6**: p. 581.
102. Cai, W., et al., *The Biosynthesis of Capuramycin-type Antibiotics: IDENTIFICATION OF THE A-102395 BIOSYNTHETIC GENE CLUSTER, MECHANISM OF SELF-RESISTANCE, AND FORMATION OF URIDINE-5'-CARBOXAMIDE.* Journal of Biological Chemistry, 2015. **290**(22): p. 13710-13724.
103. Huang, Y., et al., *Pyridoxal-5'-phosphate as an oxygenase cofactor: Discovery of a carboxamide-forming, α-amino acid monooxygenase-decarboxylase.* Proceedings of the National Academy of Sciences, 2018.
104. Funabashi, M., et al., *Identification of the biosynthetic gene cluster of A-500359s in Streptomyces griseus SANK60196.* The Journal Of Antibiotics, 2009. **62**: p. 325.
105. Yang, Z., et al., *Functional and kinetic analysis of the phosphotransferase CapP conferring selective self-resistance to capuramycin antibiotics.* J Biol Chem, 2010. **285**(17): p. 12899-905.
106. Cai, W., et al., *The Biosynthesis of Capuramycin-Type Antibiotics: Identification of the A-102395 Biosynthetic Gene Cluster, Mechanism of Self-Resistance, and Formation of Uridine-5'-*

Carboxamide. Journal of Biological Chemistry, 2015.

107. Green, J.M. and B.P. Nichols, *p-Aminobenzoate biosynthesis in Escherichia coli. Purification of aminodeoxychorismate lyase and cloning of pabC.* Journal of Biological Chemistry, 1991. **266**(20): p. 12971-12975.
108. Van Lanen, S.G., S. Lin, and B. Shen, *Biosynthesis of the enediyne antitumor antibiotic C-1027 involves a new branching point in chorismate metabolism.* Proceedings of the National Academy of Sciences, 2008. **105**(2): p. 494-499.
109. Sahr, T., et al., *Folate synthesis in plants: purification, kinetic properties, and inhibition of aminodeoxychorismate synthase.* Biochemical Journal, 2006. **396**(1): p. 157-162.
110. Ravanel, S., R. Douce, and F. Rébeillé, *The Uniqueness of Tetrahydrofolate Synthesis and One-Carbon Metabolism in Plants*, in *Plant Mitochondria: From Genome to Function*, D.A. Day, A.H. Millar, and J. Whelan, Editors. 2004, Springer Netherlands: Dordrecht. p. 277-292.
111. Douce, R., et al., *The glycine decarboxylase system: a fascinating complex.* Trends in Plant Science, 2001. **6**(4): p. 167-176.
112. Goncharoff, P. and B.P. Nichols, *Nucleotide sequence of Escherichia coli pabB indicates a common evolutionary origin of p-aminobenzoate synthetase and anthranilate synthetase.* Journal of Bacteriology, 1984. **159**(1): p. 57-62.
113. Yanofsky, C., et al., *The complete nucleotide sequence of the tryptophan operon of Escherichia coli.* Nucleic Acids Research, 1981. **9**(24): p. 6647-6668.
114. Policastro, P.P., et al., *trans-6-Amino-5-[(1-carboxyethenyl)oxy]-1,3-cyclohexadiene-1-carboxylic acid: an intermediate in the biosynthesis of anthranilate from chorismate.* Journal of the American Chemical Society, 1984. **106**(8): p. 2443-2444.
115. Old, I.G., et al., *Mapping of genes on the linear chromosome of the bacterium Borrelia burgdorferi: Possible locations for its origin of replication.* FEMS Microbiology Letters, 1992. **99**(2-3): p. 245-250.
116. Huang, M. and F. Gibson, *Biosynthesis of 4-Aminobenzoate in Escherichia coli.* Journal of Bacteriology, 1970. **102**(3): p. 767-773.
117. Kaplan, J.B., B.P. Nichols, and S. Brenner, *Nucleotide sequence of Escherichia coli pabA and its evolutionary relationship to trp (G) D.* Journal of Molecular Biology, 1983. **168**(3): p. 451-468.
118. Viswanathan, V.K., J.M. Green, and B.P. Nichols, *Kinetic characterization of 4-amino 4-deoxychorismate synthase from Escherichia coli.* Journal of Bacteriology, 1995. **177**(20): p. 5918-23.
119. Roux, B. and C.T. Walsh, *p-Aminobenzoate synthesis in Escherichia coli: kinetic and mechanistic characterization of the amidotransferase PabA.* Biochemistry, 1992. **31**(30): p. 6904-6910.
120. Ye, Q.Z., J. Liu, and C.T. Walsh, *p-Aminobenzoate synthesis in Escherichia coli: purification and characterization of PabB as aminodeoxychorismate synthase and enzyme X as aminodeoxychorismate lyase.* Proceedings of the National Academy of Sciences, 1990. **87**(23): p. 9391-9395.
121. Teng, C.Y.P. and B. Ganem, *Shikimate-derived metabolites. 13. A key intermediate in the biosynthesis of anthranilate from chorismate.* Journal of the American Chemical Society, 1984. **106**(8): p. 2463-2464.

122. Roux, B. and C.T. Walsh, *p-Aminobenzoate synthesis in Escherichia coli: Mutational analysis of three conserved amino acid residues of the amidotransferase PabA.* Biochemistry, 1993. **32**(14): p. 3763-3768.
123. Teng, C.Y.P., et al., *Total synthesis of (.+-.)-4-amino-4-deoxychorismic acid: a key intermediate in the biosynthesis of p-aminobenzoic acid and L-p-aminophenylalanine.* Journal of the American Chemical Society, 1985. **107**(17): p. 5008-5009.
124. Parsons, J.F., et al., *Structure of Escherichia coli Aminodeoxychorismate Synthase: Architectural Conservation and Diversity in Chorismate-Utilizing Enzymes.* Biochemistry, 2002. **41**(7): p. 2198-2208.
125. Morollo, A.A. and M.J. Eck, *Structure of the cooperative allosteric anthranilate synthase from Salmonella typhimurium.* Nature Structural Biology, 2001. **8**: p. 243.
126. Bulloch, E.M.M., et al., *Identification of 4-Amino-4-deoxychorismate Synthase as the Molecular Target for the Antimicrobial Action of (6S)-6-Fluoroshikimate.* Journal of the American Chemical Society, 2004. **126**(32): p. 9912-9913.
127. Klaus, S.M.J., et al., *Higher Plant Plastids and Cyanobacteria Have Folate Carriers Related to Those of Trypanosomatids.* Journal of Biological Chemistry, 2005. **280**(46): p. 38457-38463.
128. Criado, L.M., J.F. Martín, and J. Gil, *The pab gene of Streptomyces griseus, encoding p-aminobenzoic acid synthase, is located between genes possibly involved in candicidin biosynthesis.* Gene, 1993. **126**(1): p. 135-139.
129. Edman, J.C., A.L. Goldstein, and J.G. Erbe, *Para-aminobenzoate synthase gene of Saccharomyces cerevisiae encodes a bifunctional enzyme.* Yeast, 1993. **9**(6): p. 669-675.
130. Fu, J.C., L. Ding, and S. Clarke, *Purification, gene cloning, and sequence analysis of an L-isoaspartyl protein carboxyl methyltransferase from Escherichia coli.* Journal of Biological Chemistry, 1991. **266**(22): p. 14562-14572.
131. Green, J.M., W.K. Merkel, and B.P. Nichols, *Characterization and sequence of Escherichia coli pabC, the gene encoding aminodeoxychorismate lyase, a pyridoxal phosphate-containing enzyme.* Journal of Bacteriology, 1992. **174**(16): p. 5317-5323.
132. Basset, G.J.C., et al., *Folate synthesis in plants: the last step of the p-aminobenzoate branch is catalyzed by a plastidial aminodeoxychorismate lyase.* The Plant Journal, 2004. **40**(4): p. 453-461.
133. Nakai, T., et al., *Three-Dimensional Structure of 4-Amino-4-Deoxychorismate Lyase from Escherichia coli1.* The Journal of Biochemistry, 2000. **128**(1): p. 29-38.
134. Seiler, C.Y., et al., *DNASU plasmid and PSI:Biology-Materials repositories: resources to accelerate biological research.* Nucleic Acids Res, 2014. **42**(Database issue): p. D1253-60.
135. Walsh, C.T., et al., *Molecular studies on enzymes in chorismate metabolism and the enterobactin biosynthetic pathway.* Chemical Reviews, 1990. **90**(7): p. 1105-1129.
136. He, Z., et al., *Conservation of Mechanism in Three Chorismate-Utilizing Enzymes.* Journal of the American Chemical Society, 2004. **126**(8): p. 2378-2385.
137. Zhang, G., et al., *Characterization of the Amicetin Biosynthesis Gene Cluster from Streptomyces vinaceusdrappus NRRL 2363 Implicates Two Alternative Strategies for Amide Bond Formation.* Applied and Environmental Microbiology, 2012. **78**(7): p. 2393-2401.
138. Blanc, V., et al., *Identification and analysis of genes from Streptomyces pristinaespiralis*

encoding enzymes involved in the biosynthesis of the 4-dimethylamino-l-phenylalanine precursor of pristinamycin I. Molecular Microbiology, 1997. **23**(2): p. 191-202.

139. He, J. and C. Hertweck, *Iteration as Programmed Event during Polyketide Assembly; Molecular Analysis of the Aureothin Biosynthesis Gene Cluster.* Chemistry & Biology, 2003. **10**(12): p. 1225-1232.
140. Campelo, A.B. and J.A. Gil, *The candicidin gene cluster from Streptomyces griseus IMRU 3570.* Microbiology, 2002. **148**(1): p. 51-59.
141. Brown, M.P., K.A. Aidoo, and L.C. Vining, *A role for pabAB, a p-aminobenzoate synthase gene of Streptomyces venezuelae ISP5230, in chloramphenicol biosynthesis.* Microbiology, 1996. **142**(6): p. 1345-1355.
142. Cociancich, S., et al., *The gyrase inhibitor albicidin consists of p-aminobenzoic acids and cyanoalanine.* Nature Chemical Biology, 2015. **11**: p. 195.
143. Eriani, G., et al., *Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs.* Nature, 1990. **347**: p. 203.
144. Sankaranarayanan, R., et al., *The Structure of Threonyl-tRNA Synthetase-tRNAThr Complex Enlightens Its Repressor Activity and Reveals an Essential Zinc Ion in the Active Site.* Cell, 1999. **97**(3): p. 371-381.
145. Asturias, F.J., et al., *Structure and molecular organization of mammalian fatty acid synthase.* Nature Structural &Amp; Molecular Biology, 2005. **12**: p. 225.
146. Ogle, J.M., A.P. Carter, and V. Ramakrishnan, *Insights into the decoding mechanism from recent ribosome structures.* Trends Biochem Sci, 2003. **28**(5): p. 259-66.
147. Finking, R. and M.A. Marahiel, *Biosynthesis of nonribosomal peptides1.* Annu Rev Microbiol, 2004. **58**: p. 453-88.
148. Gevers, W., H. Kleinkauf, and F. Lipmann, *The activation of amino acids for biosynthesis of gramicidin S.* Proceedings of the National Academy of Sciences, 1968. **60**(1): p. 269-276.
149. Saleem, M., et al., *Marine natural products of fungal origin.* Natural Product Reports, 2007. **24**(5): p. 1142-1152.
150. Kasahara, T. and T. Kato, *A new redox-cofactor vitamin for mammals.* Nature, 2003. **422**: p. 832.
151. Marahiel, M.A., *Protein templates for the biosynthesis of peptide antibiotics.* Chem Biol, 1997. **4**(8): p. 561-7.
152. Schwarzer, D. and M.A. Marahiel, *Multimodular biocatalysts for natural product assembly.* Naturwissenschaften, 2001. **88**(3): p. 93-101.
153. von Dohren, H., et al., *Multifunctional Peptide Synthetases.* Chem Rev, 1997. **97**(7): p. 2675-2706.
154. Guenzi, E., et al., *Characterization of the Syringomycin Synthetase Gene Cluster: A LINK BETWEEN PROKARYOTIC AND EUKARYOTIC PEPTIDE SYNTHETASES.* Journal of Biological Chemistry, 1998. **273**(49): p. 32857-32863.
155. Stachelhaus, T. and M.A. Marahiel, *Modular Structure of Peptide Synthetases Revealed by Dissection of the Multifunctional Enzyme GrsA.* Journal of Biological Chemistry, 1995. **270**(11): p. 6163-6169.
156. Schwarzer, D., R. Finking, and M.A. Marahiel, *Nonribosomal peptides: from genes to products.*

Natural Product Reports, 2003. **20**(3): p. 275-287.

157. May, J.J., et al., *Crystal structure of DhbE, an archetype for aryl acid activating domains of modular nonribosomal peptide synthetases.* Proceedings of the National Academy of Sciences, 2002. **99**(19): p. 12120-12125.
158. Stachelhaus, T., H.D. Mootz, and M.A. Marahiel, *The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases.* Chemistry & Biology, 1999. **6**(8): p. 493-505.
159. Weber, T., et al., *Solution structure of PCP, a prototype for the peptidyl carrier domains of modular peptide synthetases.* Structure, 2000. **8**(4): p. 407-418.
160. Keating, T.A., C.G. Marshall, and C.T. Walsh, *Reconstitution and Characterization of the Vibrio cholerae Vibriobactin Synthetase from VibB, VibE, VibF, and VibH.* Biochemistry, 2000. **39**(50): p. 15522-15530.
161. Keating, T.A., et al., *The structure of VibH represents nonribosomal peptide synthetase condensation, cyclization and epimerization domains.* Nature Structural Biology, 2002. **9**: p. 522.
162. Dieckmann, R., et al., *Expression of an active adenylate-forming domain of peptide synthetases corresponding to acyl-CoA-synthetases.* FEBS Letters, 1995. **357**(2): p. 212-216.
163. Conti, E., et al., *Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S.* The EMBO Journal, 1997. **16**(14): p. 4174-4183.
164. Du, L., et al., *The biosynthetic gene cluster for the antitumor drug bleomycin from Streptomyces verticillus ATCC15003 supporting functional interactions between nonribosomal peptide synthetases and a polyketide synthase.* Chemistry & Biology, 2000. **7**(8): p. 623-642.
165. Onesti, S., A.D. Miller, and P. Brick, *The crystal structure of the lysyl-tRNA synthetase (LysU) from Escherichia coli.* Structure, 1995. **3**(2): p. 163-176.
166. Cusack, S., A. Yaremchuk, and M. Tukalo, *The crystal structure of the ternary complex of T.thermophilus seryl-tRNA synthetase with tRNA(Ser) and a seryl-adenylate analogue reveals a conformational switch in the active site.* The EMBO Journal, 1996. **15**(11): p. 2834-2842.
167. Baldwin, T.O., *Firefly luciferase: the structure is known, but the mystery remains.* Structure, 1996. **4**(3): p. 223-228.
168. Conti, E., N.P. Franks, and P. Brick, *Crystal structure of firefly luciferase throws light on a superfamily of adenylate-forming enzymes.* Structure, 1996. **4**(3): p. 287-298.
169. Stachelhaus, T., A. Schneider, and M. Marahiel, *Rational design of peptide antibiotics by targeted replacement of bacterial and fungal domains.* Science, 1995. **269**(5220): p. 69-72.
170. Marahiel, M.A., *Multidomain enzymes involved in peptide synthesis.* FEBS Letters, 1992. **307**(1): p. 40-43.
171. Kleinkauf, H. and H. von Döhren, *Review Nonribosomal biosynthesis of peptide antibiotics*, in *EJB Reviews 1990*, P. Christen and E. Hofmann, Editors. 1991, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 151-165.
172. Stachelhaus, T., A. Schneider, and M.A. Marahiel, *Engineered biosynthesis of peptide antibiotics.* Biochemical Pharmacology, 1996. **52**(2): p. 177-186.
173. Stachelhaus, T., A. Hüser, and M.A. Marahiel, *Biochemical characterization of peptidyl carrier protein (PCP), the thiolation domain of multifunctional peptide synthetases.* Chemistry & Biology, 1996. **3**(11): p. 913-921.

174. Stein, T., et al., *Gramicidin S synthetase 1 (phenylalanine racemase), a prototype of amino acid racemases containing the cofactor 4'-phosphopantetheine.* Biochemistry, 1995. **34**(14): p. 4633-4642.
175. Stein, T., et al., *Detection of 4'-phosphopantetheine at the thioester binding site for l-valine of gramicidinS synthetase 2.* FEBS Letters, 1994. **340**(1-2): p. 39-44.
176. Lambalot, R.H., et al., *A new enzyme superfamily — the phosphopantetheinyl transferases.* Chemistry & Biology, 1996. **3**(11): p. 923-936.
177. Pfeifer, E., et al., *Characterization of tyrocidine synthetase 1 (TY1): Requirement of posttranslational modification for peptide biosynthesis.* Biochemistry, 1995. **34**(22): p. 7450-7459.
178. von Dohren, H., R. Dieckmann, and M. Pavela-Vrancic, *The nonribosomal code.* Chem Biol, 1999. **6**(10): p. R273-9.
179. Wong, H.C., et al., *The solution structure of acyl carrier protein from Mycobacterium tuberculosis.* J Biol Chem, 2002. **277**(18): p. 15874-80.
180. Mofid, M.R., R. Finking, and M.A. Marahiel, *Recognition of Hybrid Peptidyl Carrier Proteins/Acyl Carrier Proteins in Nonribosomal Peptide Synthetase Modules by the 4'-Phophopantetheinyl Transferases AcpS and Sfp.* Journal of Biological Chemistry, 2002. **277**(19): p. 17023-17031.
181. Flaman, A.S., et al., *Site-directed mutagenesis of acyl carrier protein (ACP) reveals amino acid residues involved in ACP structure and acyl-ACP synthetase activity.* J Biol Chem, 2001. **276**(38): p. 35934-9.
182. Byers, D.M. and H. Gong, *Acyl carrier protein: structure–function relationships in a conserved multifunctional protein family.* Biochemistry and Cell Biology, 2007. **85**(6): p. 649-662.
183. Li, Q., et al., *Solution Structure and Backbone Dynamics of the Holo Form of the Frenolicin Acyl Carrier Protein.* Biochemistry, 2003. **42**(16): p. 4648-4657.
184. Linne, U. and M.A. Marahiel, *Control of directionality in nonribosomal peptide synthesis: role of the condensation domain in preventing misinitiation and timing of epimerization.* Biochemistry, 2000. **39**(34): p. 10439-47.
185. Stachelhaus, T., et al., *Peptide bond formation in nonribosomal peptide biosynthesis. Catalytic role of the condensation domain.* J Biol Chem, 1998. **273**(35): p. 22773-81.
186. MacKelvie, S.H., P.D. Andrews, and M.J. Stark, *The Saccharomyces cerevisiae gene SDS22 encodes a potential regulator of the mitotic function of yeast type 1 protein phosphatase.* Molecular and Cellular Biology, 1995. **15**(7): p. 3777-3785.
187. Keating, T.A., et al., *The structure of VibH represents nonribosomal peptide synthetase condensation, cyclization and epimerization domains.* Nat Struct Biol, 2002. **9**(7): p. 522-6.
188. Davies, C., et al., *The 1.8 A crystal structure and active-site architecture of beta-ketoacyl-acyl carrier protein synthase III (FabH) from escherichia coli.* Structure, 2000. **8**(2): p. 185-95.
189. Guenzi, E., et al., *Coordinate transcription and physical linkage of domains in surfactin synthetase are not essential for proper assembly and activity of the multienzyme complex.* J Biol Chem, 1998. **273**(23): p. 14403-10.
190. Mootz, H.D., D. Schwarzer, and M.A. Marahiel, *Construction of hybrid peptide synthetases by module and domain fusions.* Proceedings of the National Academy of Sciences, 2000. **97**(11):

p. 5848-5853.

191. D'Souza, C., M.M. Nakano, and P. Zuber, *Identification of comS, a gene of the srfA operon that regulates the establishment of genetic competence in Bacillus subtilis.* Proceedings of the National Academy of Sciences of the United States of America, 1994. **91**(20): p. 9397-9401.
192. Gutiérrez, S., et al., *Characterization of the Cephalosporium acremonium pcbAB gene encoding alpha-aminoadipyl-cysteinyl-valine synthetase, a large multidomain peptide synthetase: linkage to the pcbC gene as a cluster of early cephalosporin biosynthetic genes and evidence of multiple functional domains.* Journal of Bacteriology, 1991. **173**(7): p. 2354-2365.
193. Panaccione, D.G., *Multiple families of peptide synthetase genes from ergopeptine-producing fungi.* Mycological Research, 1996. **100**(4): p. 429-436.
194. Turgay, K., M. Krause, and M.A. Marahiel, *Four homologous domains in the primary structure of GrsB are related to domains in a superfamily of adenylate-forming enzymes.* Mol Microbiol, 1992. **6**(4): p. 529-46.
195. D'Souza, C., et al., *Translation of the open reading frame encoded by comS, a gene of the srf operon, is necessary for the development of genetic competence, but not surfactin biosynthesis, in Bacillus subtilis.* Journal of Bacteriology, 1995. **177**(14): p. 4144-4148.
196. Lin, G.-H., et al., *Molecular Cloning and Characterization of Fengycin Synthetase Gene fenB from Bacillus subtilis.* Journal of Bacteriology, 1998. **180**(5): p. 1338-1341.
197. Schmidt, M.F., M.R. Groves, and J. Rademann, *Dynamic Substrate Enhancement for the Identification of Specific, Second-Site-Binding Fragments Targeting a Set of Protein Tyrosine Phosphatases.* ChemBioChem, 2011. **12**(17): p. 2640-2646.
198. Timothy J. Manning , Z.L.W., *Advertorial: Introduction of a Convenient, Non-Radioactive Universal Kinase Activity Assay Suitable for High-Throughput Screening.* Genetic Engineering & Biotechology News, 2012. **32**(21).
199. Belecki, K. and C.A. Townsend, *Biochemical Determination of Enzyme-Bound Metabolites: Preferential Accumulation of a Programmed Octaketide on the Enediyne Polyketide Synthase CalE8.* Journal of the American Chemical Society, 2013. **135**(38): p. 14339-14348.
200. Cronan, J.E. and A.L. Klages, *Chemical synthesis of acyl thioesters of acyl carrier protein with native structure.* Proceedings of the National Academy of Sciences, 1981. **78**(9): p. 5440-5444.
201. Keller, U., et al., *The Actinomycin Biosynthetic Gene Cluster of Streptomyces chrysomallus: a Genetic Hall of Mirrors for Synthesis of a Molecule with Mirror Symmetry.* Journal of Bacteriology, 2010. **192**(10): p. 2583-2595.
202. Staunton, J. and K.J. Weissman, *Polyketide biosynthesis: a millennium review.* Natural Product Reports, 2001. **18**(4): p. 380-416.
203. Shen, B., *Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms.* Current Opinion in Chemical Biology, 2003. **7**(2): p. 285-295.
204. Staunton, J., *Combinatorial biosynthesis of erythromycin and complex polyketides.* Current Opinion in Chemical Biology, 1998. **2**(3): p. 339-345.
205. Leadlay, P.F., *Combinatorial approaches to polyketide biosynthesis.* Current Opinion in Chemical Biology, 1997. **1**(2): p. 162-168.

206. Matharu, A.-L., et al., *MCAT is not required for in vitro polyketide synthesis in a minimal actinorhodin polyketide synthase from Streptomyces coelicolor.* Chemistry & Biology, 1998. **5**(12): p. 699-711.
207. Crosby, J., et al., *Acylation of Streptomyces type II polyketide synthase acyl carrier proteins.* FEBS Letters, 1998. **433**(1-2): p. 132-138.
208. Cox, R.J., et al., *Post-translational modification of heterologously expressed Streptomyces type II polyketide synthase acyl carrier proteins.* FEBS Letters, 1997. **405**(3): p. 267-272.
209. HOLAK, T.A., et al., *Three-dimensional structure of acyl carrier protein in solution determined by nuclear magnetic resonance and the combined use of dynamical simulated annealing and distance geometry.* European Journal of Biochemistry, 1988. **175**(1): p. 9-15.
210. Andrec, M., R.B. Hill, and J.H. Prestegard, *Amide exchange rates in Escherichia coli acyl carrier protein: Correlation with protein structure and dynamics.* Protein Science, 1995. **4**(5): p. 983-993.
211. Hitchman, T.S., et al., *Catalytic self-acylation of type II polyketide synthase acyl carrier proteins.* Chemistry & Biology, 1998. **5**(1): p. 35-47.
212. Zhou, P., G. Florova, and K.A. Reynolds, *Polyketide synthase acyl carrier protein (ACP) as a substrate and a catalyst for malonyl ACP biosynthesis.* Chemistry & Biology, 1999. **6**(8): p. 577-584.
213. Brahmachari, G., *Chemistry and pharmacology of naturally occurring bioactive compounds*. 2013: CRC Press.
214. Bisang, C., et al., *A chain initiation factor common to both modular and aromatic polyketide synthases.* Nature, 1999. **401**: p. 502.
215. Fujii, I., et al., *Identification of Claisen cyclase domain in fungal polyketide synthase WA, a naphthopyrone synthase of Aspergillus nidulans.* Chemistry & Biology, 2001. **8**(2): p. 189-197.
216. Keatinge-Clay, A.T., *Stereocontrol within polyketide assembly lines.* Natural product reports, 2016. **33**(2): p. 141-149.
217. Keatinge-Clay, A.T. and R.M. Stroud, *The structure of a ketoreductase determines the organization of the β-carbon processing enzymes of modular polyketide synthases.* Structure, 2006. **14**(4): p. 737-748.
218. Rawlings, M. and J.E. Cronan, *The gene encoding Escherichia coli acyl carrier protein lies within a cluster of fatty acid biosynthetic genes.* Journal of Biological Chemistry, 1992. **267**(9): p. 5751-4.
219. Toomey, R.E. and S.J. Wakil, *Studies on the mechanism of fatty acid synthesis XV. Preparation and general properties of β-ketoacyl acyl carrier protein reductase from Escherichia coli.* Biochimica et Biophysica Acta (BBA) - Lipids and Lipid Metabolism, 1966. **116**(2): p. 189-197.
220. Lai, C.-Y. and J.E. Cronan, *Isolation and Characterization of β-Ketoacyl-Acyl Carrier Protein Reductase (fabG) Mutants of Escherichia coli and Salmonella enterica Serovar Typhimurium.* Journal of Bacteriology, 2004. **186**(6): p. 1869-1878.
221. Zhang, Y.-M. and C.O. Rock, *Evaluation of Epigallocatechin Gallate and Related Plant Polyphenols as Inhibitors of the FabG and FabI Reductases of Bacterial Type II Fatty-acid Synthase.* Journal of Biological Chemistry, 2004. **279**(30): p. 30994-31001.
222. White, S.W., et al., *THE STRUCTURAL BIOLOGY OF TYPE II FATTY ACID BIOSYNTHESIS.* Annual

Review of Biochemistry, 2005. **74**(1): p. 791-831.

223. Price, A.C., et al., *Cofactor-Induced Conformational Rearrangements Establish a Catalytically Competent Active Site and a Proton Relay Conduit in FabG.* Structure, 2004. **12**(3): p. 417-428.

224. Palaniappan, N., et al., *Production of hygromycin A analogs in Streptomyces hygroscopicus NRRL 2388 through identification and manipulation of the biosynthetic gene cluster.* Chem Biol, 2006. **13**(7): p. 753-64.

Vita

Erfu Yan

Education

Aug 2013-Present Ph.D. candidate in Pharmaceutical Science, College of Pharmacy, University of Kentucky

Sept 2010-Jun 2013 M.S., Microbial and Biochemical Pharmacy, Tianjin University, China

Sept 2006-Jun 2010 B.E., Pharmaceutical Engineering, Department of Chemical Engineering, Tianjin University, China

Professionals

Research Assistant, College of Pharmacy, University of Kentucky, KY

(2013 - 2018)

Publications

[1] Ji Li, Zhipeng Tang, Rui Hu, Qiang Fu, Erfu Yan, Shaoying Wang. Probing surface hydrophobicity of individual protein at single-molecule resolution using solid-state nanopores, Science China Materials, 58(2015), pp: 455-466

[2] Hongji Zhu, Kai Sheng, Erfu Yan, Jianjun Qiao*, Feng Lv, Extraction, purification and antibacterial activities of a polysaccharide from spent mushroom substrate. International Journal of Biological Macromolecules, 50(2012), pp: 840-843.

[3] Erfu Yan, Lian Xu, Wenjuan Liu, Lijun Zhou*, Novel anticancer drug molecular target STAT3 and its advancement. Progress in Modern Biomedicine 07-049(2013)

[4] Yang Wang, Chaoyang Deng, Lu Yang, Erfu Yan, Tao Guo, Yanmin Li, Marvin Xuejun Xu*,

Mechanism of the inhibition of the STAT3 signaling pathway by EGCG, 30(2013), pp: 2691-2696.

[5] Shaoying Wang, Erfu Yan, Farzin Haque, Peixuan Guo*, Three-step channel conformational changes common to DNA translocases of bacterial viruses T3, T4, SPP1, and Phi29. Virology, pii: S0042-6822(16)30073-3. DOI: 10.1016/j.virol.2016.04.015. [Epub ahead of print]

[6] Li Tian, Jiaheng Liu, Li Ma, Lei Zhang, Erfu Yan and Hongji Zhu*, Isolation and purification of antioxidant and ACE-inhibitory peptides from yak (bos grunniens) skin, Journal of Food Processing and Preservation, DOI: 10.1111/jfpp.13123. [Epub ahead of print]

[7] Erfu Yan, Shaoying Wang, Farzin Haque, Peixuan Guo*, Alteration of channel chirality to control direction of DNA translocation, in preparation.

[8] Erfu Yan, Zheng Cui, Steven Van Lanen*, The biosynthesis of capuramycin-type antibiotics: polyamide formation and characterization of the A-102395 biosynthetic gene cluster, in preparation.

[9] Erfu Yan, Zheng Cui, Steven Van Lanen*, Identification and Characterization of the stereochemitry of capuramycin of A-102395 in Streptomyces, in preparation.

[10] Erfu Yan, Zheng Cui, Steven Van Lanen*, type II polyketide synthase systems recruit fatty acid biosynthetic machinery during polyketide assembly in the biosynthesis pathway of capuramycin of A-102395 in Streptomyces, in preparation.

Erfu Yan